

---

# MEGI

**MESTRADO**

Estatística e Gestão de Informação

---

---

## **Utilização de redes para a detecção de casos de fraude em apólices de seguro automóvel.**

Caso de estudo em seguradoras portuguesas.

---

Carlos Miguel Fernandes Francisco

---

Trabalho de Projecto apresentado como requisito parcial  
para obtenção do grau de Mestre em Estatística e Gestão de  
Informação

Instituto Superior de Estatística e Gestão de Informação  
Universidade Nova de Lisboa

**Utilização de redes para a detecção de casos de fraude em apólices  
de seguro automóvel.**

Caso de estudo em seguradoras portuguesas.

Por:

Carlos Miguel Fernandes Francisco

Trabalho de Projecto apresentado como requisito parcial para obtenção do grau de  
Mestre em Estatística e Gestão de Informação

Orientador: Prof. Doutor Rui Gonçalves

Novembro 2014

## **Resumo e Palavras-Chave**

O tema da fraude nas seguradoras tem sido um objecto de estudo, durante as recentes décadas, por investigadores de diferentes áreas da sociedade, tanto do meio académico como por especialistas da indústria, reflectindo assim a importância deste tema no campo da economia e da aceitação social.

O principal objectivo teórico é a criação de uma rede que revele as relações entre uma base de dados de dados de sinistros segmentada, sendo esta providenciada por três grandes seguradoras que operam no mercado dos seguros não-vida português, obtendo assim informação sobre as características do cliente fraudulento.

Mais especificamente, os segmentos da análise são feitos através da análise de duas variáveis alvo presentes nesta que dão informação se o cliente é não-suspeito e não-confirmado, não-suspeito e confirmado, suspeito e não confirmado e suspeito e confirmado. Após a análise, os quatro segmentos são criados e posteriormente, a rede é elaborada com base na relação das diferentes características dos segmentos criados.

Os resultados mostram que os segmentos que se referem aos casos suspeitos e não confirmados e os casos suspeitos e confirmados são os que mais se relacionam, por outro lado, o segmento dos casos não suspeitos e não confirmados é o que menos características em comum.

A presente dissertação será escrita de acordo com o antigo acordo ortográfico.

**Palavras-chave:** Fraude nas seguradoras, Redes, Apólices, Relações, Segmentos,

## ***Abstract***

The insurance fraud theme has been a subject of a various number of studies, during the past decades, by researchers of different areas of the modern society, from both academic and industry experts reflecting there by the importance of this thematic in the fields of economics and social acceptance.

The main theoretical objective is the creation of a network that reveals the relationships between a segmented insurance claims sample that has been provided by three major insurance companies operating in the Portuguese non-life insurance market, providing therefore information about the characteristics of fraudulent customer.

More specifically, the sample segmentations are made by analyzing two targets variables that are contained with it and give information whether the claim is non-suspect and non-confirmed, non-suspect and confirmed, suspect and non-confirmed and for the last case suspect and confirmed. After the analysis, the four segments are created and then the relationships of the different characteristics of the segments are created, giving birth to the network of segments.

The results show that the segment referring the claims that are suspect and non-confirmed and the suspect and confirmed are the most related segments, on the other hand, the segment of the non-suspect and non-confirmed are the ones that have less characteristics in common.

**Key Words: Insurance Fraud, Network, Claims, Relationships, Segments.**

**Jel Codes: G22, L14**

# Índice

I.	Introdução.....	5
a)	Objectivos do Estudo e Enquadramento do problema.....	5
b)	Importância e Justificação do Problema.....	6
II.	Revisão Literária.....	8
a)	Fraude nos seguros.....	8
a.	Contextualização.....	8
b.	Enquadramento histórico.....	14
c.	Compreensão jurídica.....	15
b)	A fraude no seguro automóvel.....	17
c)	Redes.....	20
a.	As ciências da complexidade.....	20
b.	Contextualização histórica.....	21
d)	Tipologias de redes.....	23
III.	Metodologia.....	28
a)	Descrição da Investigação.....	28
IV.	Resultados.....	31
a)	Segmentação e análise da amostra.....	31
b)	Criação de matriz de semelhança.....	34
V.	Conclusões.....	37
VI.	Limitações e investigação futura.....	39
VII.	Bibliografia.....	41
VIII.	Anexos.....	44
a)	Gráficos.....	44
b)	Tabelas.....	70
c)	Matrizes.....	98

## ÍNDICE DE FIGURAS

Figura 1 – Aplicação de técnicas antifraude para os diferentes casos de fraude nas seguradoras.	19
Figura 2 - As sete pontes de Königsberg e a origem da Teoria dos Grafos.	21
Figura 3 - Ilustração da rede.	22
Figura 4 – Descrição visual do processo de rewiring.	25
Figura 5 – Materialização da Rede entre segmentos.	37

## ÍNDICE DE FORMULAS

Formula 1: Coeficiente de agregação genérico	24
Formula 2 – Distancia média mais curta.	25
Formula 3 - Grau médio da rede.	26
Formula 4 – Coeficiente de agregação das redes small world.	26
Formula 5 - Coeficiente $R(g_i)$ .	26

## ÍNDICE DE TABELAS

Tabela 1 – Enumeração e descrição das 39 variáveis em utilização.	29
Tabela 2 – Estatísticas descritivas para as variáveis numéricas nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	70
Tabela 3 – Tabela de frequência para a variável Perc_Policy_Exposure nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	71
Tabela 4 – Tabela de frequência para a variável COD_Claim_Time nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	71
Tabela 5 – Tabela de frequência para a variável COD_Claim_Loc_Nuts nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	71
Tabela 6 – Tabela de frequência para a variável COD_NAT_CI nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	72
Tabela 7 – Tabela de frequência para a variável COD_Gender_CI nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	72
Tabela 8 – Tabela de frequência para a variável COD_Occupation_CI nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	72
Tabela 9 – Tabela de frequência para a variável COD_INJ_Nature_CI nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	72
Tabela 10 – Tabela de frequência para a variável COD_VI_MM nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	72
Tabela 11 – Tabela de frequência para a variável COD_VI_TYPE nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	74
Tabela 12 – Tabela de frequência para a variável COD_V1_MM nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	74
Tabela 13 – Tabela de frequência para a variável COD_V1_TYPE nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0	76
Tabela 14 – Tabela de frequência para a variável COD_V2_MM nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0	77
Tabela 15 – Tabela de frequência para a variável COD_V2_TYPE nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0	78
Tabela 16 – Tabela de frequência para a variável FLG_Claim_Wkend nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0	78



Tabela 17 – Tabela de frequência para a variável FLG_Claim_Overseas nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0	78
Tabela 18 – Tabela de frequência para a variável FLG_Claim_Insured nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0	78
Tabela 19 – Tabela de frequência para a variável FLG_Claim_Police_Rep nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0	79
Tabela 20 – Tabela de frequência para a variável FLG_Claim_Injured nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0	79
Tabela 21 – Tabela de frequência para a variável FLG_Claim_Witness nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0	79
Tabela 22 – Tabela de frequência para a variável FLG_PC_Providers_CI nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0	79
Tabela 23 – Tabela de frequência para a variável FLG_VI_Chose_Garage nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0	79
Tabela 24 – Tabela de frequência para a variável FLG_VI_Repair_PFraud nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0	79
Tabela 25 – Tabela de frequência para a variável FLG_PC_V1_Provider nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0	80
Tabela 26 – Tabela de estatísticas descritivas para a variável VAL_Policy_Exposure nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0	80
Tabela 27 – Estatísticas descritivas para as variáveis numéricas nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	80
Tabela 28 – Tabela de frequência para a variável Perc_Policy_Exposure nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	80
Tabela 29 – Tabela de frequência para a variável COD_Claim_Time nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	80
Tabela 30 – Tabela de frequência para a variável COD_Claim_Loc_Nuts nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	80
Tabela 31 – Tabela de frequência para a variável COD_NAT_CI nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	81
Tabela 32 – Tabela de frequência para a variável COD_Gender_CI nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	81
Tabela 33 – Tabela de frequência para a variável COD_Occupation_CI nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	81

Tabela 34 – Tabela de frequência para a variável COD_INJ_Nature_CI nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	82
Tabela 35 – Tabela de frequência para a variável COD_VI_MM nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	82
Tabela 36 – Tabela de frequência para a variável COD_VI_TYPE nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	83
Tabela 37 – Tabela de frequência para a variável COD_V1_MM nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	83
Tabela 38 – Tabela de frequência para a variável COD_V1_TYPE nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1	83
Tabela 39 – Tabela de frequência para a variável COD_V2_MM nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1	84
Tabela 40 – Tabela de frequência para a variável COD_V2_TYPE nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1	84
Tabela 41 – Tabela de frequência para a variável FLG_Claim_Wkend nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1	84
Tabela 42 – Tabela de frequência para a variável FLG_Claim_Overseas nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1	84
Tabela 43 – Tabela de frequência para a variável FLG_Claim_Insured nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1	85
Tabela 44 – Tabela de frequência para a variável FLG_Claim_Police_Rep nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1	85
Tabela 45 – Tabela de frequência para a variável FLG_Claim_Injured nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1	85
Tabela 46 – Tabela de frequência para a variável FLG_Claim_Witness nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1	85
Tabela 47 – Tabela de frequência para a variável FLG_PC_Providers_CI nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1	85
Tabela 48 – Tabela de frequência para a variável FLG_VI_Chose_Garage nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1	85
Tabela 49 – Tabela de frequência para a variável FLG_VI_Repair_PFraud nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1	85
Tabela 50 – Tabela de frequência para a variável FLG_PC_V1_Provider nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1	85

Tabela 51 – Tabela de estatísticas descritivas para a variável VAL_Policy_Exposure nos casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1	86
Tabela 52 – Estatísticas descritivas para as variáveis numéricas nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	86
Tabela 53 – Tabela de frequência para a variável Perc_Policy_Exposure nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	86
Tabela 54 – Tabela de frequência para a variável COD_Claim_Time nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	86
Tabela 55 – Tabela de frequência para a variável COD_Claim_Loc_Nuts nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	87
Tabela 56 – Tabela de frequência para a variável COD_NAT_CI nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	87
Tabela 57 – Tabela de frequência para a variável COD_Gender_CI nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	87
Tabela 58 – Tabela de frequência para a variável COD_Occupation_CI nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	87
Tabela 59 – Tabela de frequência para a variável COD_INJ_Nature_CI nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	87
Tabela 60 – Tabela de frequência para a variável COD_VI_MM nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	87
Tabela 61 – Tabela de frequência para a variável COD_VI_TYPE nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	88
Tabela 62 – Tabela de frequência para a variável COD_V1_MM nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	88
Tabela 63 – Tabela de frequência para a variável COD_V1_TYPE nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	88
Tabela 64 – Tabela de frequência para a variável COD_V2_MM nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	88
Tabela 65 – Tabela de frequência para a variável COD_V2_TYPE nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	88
Tabela 66 – Tabela de frequência para a variável FLG_Claim_Wkend nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	89
Tabela 67 – Tabela de frequência para a variável FLG_Claim_Overseas nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	89

Tabela 68 – Tabela de frequência para a variável FLG_Claim_Insured nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	89
Tabela 69 – Tabela de frequência para a variável FLG_Claim_Police_Rep nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	89
Tabela 70 – Tabela de frequência para a variável FLG_Claim_Injured nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	89
Tabela 71 – Tabela de frequência para a variável FLG_Claim_Witness nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	89
Tabela 72 – Tabela de frequência para a variável FLG_PC_Providers_CI nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	89
Tabela 73 – Tabela de frequência para a variável FLG_VI_Chose_Garage nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	90
Tabela 74 – Tabela de frequência para a variável FLG_VI_Repair_PFraud nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	90
Tabela 75 – Tabela de frequência para a variável FLG_PC_V1_Provider nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	90
Tabela 76 – Tabela de estatísticas descritivas para a variável VAL_Policy_Exposure nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	90
Tabela 77 – Estatísticas descritivas para as variáveis numéricas nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	90
Tabela 78 – Tabela de frequência para a variável Perc_Policy_Exposure nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	90
Tabela 79 – Tabela de frequência para a variável COD_Claim_Time nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	91
Tabela 80 – Tabela de frequência para a variável COD_Claim_Loc_Nuts nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	91
Tabela 81 – Tabela de frequência para a variável COD_NAT_CI nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	91
Tabela 82 – Tabela de frequência para a variável COD_Gender_CI nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	91
Tabela 83 – Tabela de frequência para a variável COD_Occupation_CI nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	91
Tabela 84 – Tabela de frequência para a variável COD_INJ_Nature_CI nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	92

Tabela 85 – Tabela de frequência para a variável COD_VI_MM nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	92
Tabela 86 – Tabela de frequência para a variável COD_VI_TYPE nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	92
Tabela 87 – Tabela de frequência para a variável COD_V1_MM nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	93
Tabela 88 – Tabela de frequência para a variável COD_V1_TYPE nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	93
Tabela 89 – Tabela de frequência para a variável COD_V2_MM nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	93
Tabela 90 – Tabela de frequência para a variável COD_V2_TYPE nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	93
Tabela 91 – Tabela de frequência para a variável FLG_Claim_Wkend nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	93
Tabela 92 – Tabela de frequência para a variável FLG_Claim_Overseas nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	93
Tabela 93 – Tabela de frequência para a variável FLG_Claim_Insured nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	93
Tabela 94 – Tabela de frequência para a variável FLG_Claim_Police_Rep nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	93
Tabela 95 – Tabela de frequência para a variável FLG_Claim_Injured nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	94
Tabela 96 – Tabela de frequência para a variável FLG_Claim_Witness nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	94
Tabela 97 – Tabela de frequência para a variável FLG_PC_Providers_CI nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	94
Tabela 98 – Tabela de frequência para a variável FLG_VI_Chose_Garage nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	94
Tabela 99 – Tabela de frequência para a variável FLG_VI_Repair_PFraud nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	94
Tabela 100 – Tabela de frequência para a variável FLG_PC_V1_Provider nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	94
Tabela 101 – Tabela de estatísticas descritivas para a variável VAL_Policy_Exposure nos casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	94
Tabela 102 – Tabela de valores omissos para o total da amostra	94

Tabela 103 – Tabela de valores omissos para o segmento 1	95
Tabela 104 – Tabela de valores omissos para o segmento 2	96
Tabela 105 – Tabela de valores omissos para o segmento 3	96
Tabela 106 – Tabela de valores omissos para o segmento 4	97

## ÍNDICE DE GRÁFICOS

Gráfico 1 – Localização das apólices consoante o quartil de percentagem do valor seguro pela apólice para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	44
Gráfico 2 – Distribuição temporal da ocorrência dos sinistros para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	44
Gráfico 3 – Distribuição geográfica por Nut3 da ocorrência dos sinistros para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	45
Gráfico 4 – Nacionalidade do sinistrado para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	45
Gráfico 5 – Género do sinistrado para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	45
Gráfico 6 – Tipos de lesão decorrente do sinistro para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	46
Gráfico 7 – Tipos de veículo seguro para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	46
Gráfico 8 – Tipo do veículo 1 no sinistro para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	46
Gráfico 9 – Tipo do veículo 2 no sinistro para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	47
Gráfico 10 – Variável binária indicativa se o acidente ocorreu num dia de semana (0) ou se ocorreu num fim-de-semana/feriado (1) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	47
Gráfico 11 – Variável binária indicativa se o acidente ocorreu num em Portugal (0) ou se ocorreu no estrangeiro (1) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	47
Gráfico 12 – Variável binária indicativa se o tomador da apólice é reclamante (1) ou não (0) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	48
Gráfico 13 – Variável binária indicativa se existiu intervenção das autoridades (1) ou não (0) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	48
Gráfico 14 – Variável binária indicativa se existiu danos corporais (1) ou não (0) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	48
Gráfico 15 – Variável binária indicativa se existiu testemunhas do sinistro (1) ou não (0) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	49
Gráfico 16 – Variável binária que indica se há Sinistros anteriores envolvendo o Segurado e Oficina (1) ou não (0) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	49
Gráfico 17 – Variável binária que indica se o Segurado escolheu a Oficina (0) ou	49

não (1) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	
Gráfico 18 – Variável binária que indica se a Oficina está associada a fraude (1) ou não (0) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	50
Gráfico 19 – Variável binária se existem sinistros anteriores envolvendo (1) ou não (0) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	50
Gráfico 20 – Distribuição do valor seguro pela apólice para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 0.	50
Gráfico 21 – Distribuição da percentagem do valor seguro para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	51
Gráfico 22 – Distribuição temporal da ocorrência dos sinistros para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	51
Gráfico 23 – Distribuição geográfica por Nut3 da ocorrência dos sinistros para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	51
Gráfico 24 – Nacionalidade do sinistrado para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	52
Gráfico 25 – Género do sinistrado para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	52
Gráfico 26 – Tipos de lesão decorrente do sinistro para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	52
Gráfico 27 – Tipos de veiculo seguro para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	53
Gráfico 28 – Tipo do veiculo 1 no sinistro para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	53
Gráfico 29 – Tipo do veiculo 2 no sinistro para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	53
Gráfico 30 – Variável binária indicativa se o acidente ocorreu num dia de semana (0) ou se ocorreu num fim-de-semana/feriado (1) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	54
Gráfico 31 – Variável binária indicativa se o acidente ocorreu num em Portugal (0) ou se ocorreu no estrangeiro (1) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	54
Gráfico 32 – Variável binária indicativa se o tomador da apólice é reclamante (1) ou não (0) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	54
Gráfico 33 – Variável binária indicativa se existiu intervenção das autoridades (1) ou não (0) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	55
Gráfico 34 – Variável binária indicativa se existiu danos corporais (1) ou não (0) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	55
Gráfico 35 – Variável binária indicativa se existiu testemunhas do sinistro (1) ou	55



não (0) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	
Gráfico 36 – Variável binária que indica se há Sinistros anteriores envolvendo o Segurado e Oficina (1) ou não (0) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	56
Gráfico 37 – Variável binária que indica se o Segurado escolheu a Oficina (0) ou não (1) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	56
Gráfico 38 – Variável binária que indica se a Oficina está associada a fraude (1) ou não (0) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	56
Gráfico 39 – Variável binária se existem sinistros anteriores envolvendo (1) ou não (0) para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	57
Gráfico 40 – Distribuição do valor seguro pela apólice para os casos onde FLG_Susp_Fraud = 0 e FLG_Conf_Fraud = 1.	57
Gráfico 41 – Localização das apólices consoante o quartil de percentagem do valor seguro pela apólice para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	57
Gráfico 42 – Distribuição temporal da ocorrência dos sinistros para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	58
Gráfico 43 – Distribuição geográfica por Nut3 da ocorrência dos sinistros para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	58
Gráfico 44 – Nacionalidade do sinistrado para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	58
Gráfico 45 – Género do sinistrado para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	59
Gráfico 46 – Tipos de lesão decorrente do sinistro para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	59
Gráfico 47 – Tipos de veículo seguro para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	59
Gráfico 48 – Tipo do veículo 1 no sinistro para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	60
Gráfico 49 – Tipo do veículo 2 no sinistro para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	60
Gráfico 50 – Variável binária indicativa se o acidente ocorreu num dia de semana (0) ou se ocorreu num fim-de-semana/feriado (1) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	60
Gráfico 51 – Variável binária indicativa se o acidente ocorreu num em Portugal (0) ou se ocorreu no estrangeiro (1) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	61
Gráfico 52 – Variável binária indicativa se o tomador da apólice é reclamante (1)	61

ou não (0) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	
Gráfico 53 – Variável binária indicativa se existiu intervenção das autoridades (1) ou não (0) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	61
Gráfico 54 – Variável binária indicativa se existiu danos corporais (1) ou não (0) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	62
Gráfico 55 – Variável binária indicativa se existiu testemunhas do sinistro (1) ou não (0) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	62
Gráfico 56 – Variável binária que indica se há Sinistros anteriores envolvendo o Segurado e Oficina (1) ou não (0) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	62
Gráfico 57 – Variável binária que indica se o Segurado escolheu a Oficina (0) ou não (1) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	63
Gráfico 58 – Variável binária que indica se a Oficina está associada a fraude (1) ou não (0) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	63
Gráfico 59 – Variável binária se existem sinistros anteriores envolvendo (1) ou não (0) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	63
Gráfico 60 – Distribuição do valor seguro pela apólice para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 0.	64
Gráfico 61 – Localização das apólices consoante o quartil de percentagem do valor seguro pela apólice para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	64
Gráfico 62 – Distribuição temporal da ocorrência dos sinistros para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	64
Gráfico 63 – Distribuição geográfica por Nut3 da ocorrência dos sinistros para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	65
Gráfico 64 – Nacionalidade do sinistrado para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	65
Gráfico 65 – Género do sinistrado para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	65
Gráfico 66 – Tipos de veículo seguro para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	66
Gráfico 67 – Tipo do veículo 1 no sinistro para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	66
Gráfico 68 – Tipo do veículo 2 no sinistro para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	66
Gráfico 69 – Variável binária indicativa se o acidente ocorreu num dia de semana (0) ou se ocorreu num fim-de-semana/feriado (1) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	67
Gráfico 70 – Variável binária indicativa se o acidente ocorreu num em Portugal (0)	67

ou se ocorreu no estrangeiro (1) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	
Gráfico 71 – Variável binária indicativa se o tomador da apólice é reclamante (1) ou não (0) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	67
Gráfico 72 – Variável binária indicativa se existiu intervenção das autoridades (1) ou não (0) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	68
Gráfico 73 – Variável binária indicativa se existiu danos corporais (1) ou não (0) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	68
Gráfico 74 – Variável binária indicativa se existiu testemunhas do sinistro (1) ou não (0) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	68
Gráfico 75 – Variável binária que indica se há Sinistros anteriores envolvendo o Segurado e Oficina (1) ou não (0) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	69
Gráfico 76 – Variável binária que indica se o Segurado escolheu a Oficina (0) ou não (1) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	69
Gráfico 77 – Variável binária que indica se a Oficina está associada a fraude (1) ou não (0) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	69
Gráfico 79 – Variável binária se existem sinistros anteriores envolvendo (1) ou não (0) para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	70
Gráfico 80 – Distribuição do valor seguro pela apólice para os casos onde FLG_Susp_Fraud = 1 e FLG_Conf_Fraud = 1.	70

## ÍNDICE DE MATRIZES

Matriz 1 – Relação da variável NUM_Antig_Act_Policy entre segmentos.	34
Matriz 2 – Relação da variável NUM_Antig_Alt_Policy entre segmentos.	35
Matriz 3 – Relação da variável VAL_Prop_Exposure entre segmentos	35
Matriz 4 – Relação da variável COD_Claim_Time entre segmentos.	35
Matriz 5 – Relação da variável NUM_V1_Age_FR entre segmentos.	35
Matriz 6 – Relação da variável COD_NAT_CI entre segmentos.	36
Matriz 7 – Percentagem das relações mais frequentes entre segmentos.	36
Matriz 8 – Relação da variável FLG_Claim_Wkend entre segmentos, sendo mais comum o facto de o acidente ocorrer durante a semana.	98
Matriz 9 – Relação da variável FLG_Claim_Wkend entre segmentos, sendo mais comum o facto de os sinistros ocorrerem principalmente no grande Porto e grande Lisboa.	98
Matriz 10 – Relação da variável FLG_Claim_Overseas entre segmentos, sendo mais comum o facto de os sinistros em território português.	98
Matriz 10 – Relação da variável FLG_Claim_Insured entre segmentos, sendo mais comum o facto de os sinistros não ter sido reclamado pelo tomador do seguro.	98
Matriz 11 – Relação da variável FLG_Claim_Police_Rep entre segmentos, sendo mais comum o facto de os sinistros não terem obtido a presença das autoridades.	98
Matriz 12 – Relação da variável FLG_Claim_Injured entre segmentos, sendo mais comum o facto de os sinistros não terem ocorrido lesões corporais decorrentes do mesmo.	98
Matriz 13 – Relação da variável COD_Gender_CI entre segmentos, sendo mais comum o facto de os tomadores do seguro em questão serem do género masculino.	99
Matriz 14 – Relação da variável FLG_VI_Chose_Garage entre segmentos, sendo mais comum o facto de os tomadores do seguro terem de sua vontade escolhido a oficina para as reparações dos seus veículos.	99
Matriz 15 – Relação da variável FLG_VI_Repair_PFraud entre segmentos, sendo mais comum o facto as oficinas seleccionadas pelos tomadores de seguros não terem sido identificadas por fraude anteriormente.	99
Matriz 16 – Relação da variável FLG_PC_V1_Provider entre segmentos, sendo mais comum o facto as oficinas seleccionadas pelos tomadores de seguros não terem sido seleccionadas em casos anteriores.	99



# **I. Introdução**

## **a) Objectivos do Estudo e Enquadramento do problema**

No final da década de 1980, o problema da fraude nas seguradoras, seja no ramo vida ou não vida, tornou-se um factor de interesse em todo o mundo, não só devido ao aumento da carteira de produtos disponibilizados por parte das seguradoras (para fazer face às mudanças na sociedade contemporânea<sup>1</sup>) mas principalmente devido ao aumento exponencial de danos/roubos no sector automóvel, como por exemplo o ateamento da própria viatura em chamas ou o exagero de danos causados nos bens assegurados, o que levou a alertas sobre quais das mesmas situações seriam verdadeiras ou não (Derrig 2002; Niemi 1995). Apesar disto, as seguradoras vítimas deste crime mantiveram a sua política/a prática inalterada durante a década de 90, uma vez que acreditavam que conseguiam passar o custo destas operações fraudulentas para as carteiras de apólices dos seus clientes, através do aumento dos prémios pedidos sem que isto afectasse a vantagem competitiva no mercado segurador. A competição económica entre seguradoras tornou mais difícil um acordo que tivesse como objectivo acções e campanhas de sensibilização a tomar para prevenir e combater a fraude (Viaene & Didene, 2004).

Com a crescente evolução do poder computacional ao longo das últimas décadas e consequente desenvolvimento exponencial das bases de dados corporativas a matéria-prima para a evolução do estudo das redes está a atingir um estado de maturação, constituindo assim um importante campo tanto na análise relacional entre os intervenientes do processo de seguros, como também na capacidade de tomar decisões mais precisas e com impacto relevante. Deste modo, todas as técnicas que permitam uma análise mais aprofundada de todos estes casos e que providenciem bases, ou modelos analíticos, com novas características mais robustas, para “delapidação” das base de dados das companhias de seguro são necessárias para estudar este fenómeno já de si complexo, mas que ganha hoje contornos, seja devido a factores externos às seguradoras como os

---

<sup>1</sup> Mudanças tipificadas pela alteração dos padrões de consumo, visando o aumento possessão de bens onde vigoravam principalmente automóvel individual e os bens de consumo mas também actividades financeiras como os créditos ao consumo ou à habitação.

factores sociais ou económicos, quer sejam estes tanto a nível nacional como internacional.

Posto isto, este projecto tem como objectivo principal o estudo das relações entre diferentes apólices, claramente identificadas e onde foi verificada a existência de praticas fraudulentas e/o suspeita de realização destas, de modo a que seja possível a criação de uma rede tendo como base uma amostra com dados reais proveniente de três seguradoras que operam no mercado português, no ramo de seguros não vida – mais precisamente no ramo automóvel.

Especificamente, os objectivos passam por utilizar estatísticas descritivas e de frequência para descrever as variáveis que foram seleccionadas para a análise da amostra de forma a obter uma ideia de como estão caracterizados cada segmento. Posteriormente, identificar características semelhantes entre os segmentos de forma a dar origem à rede de relações entre estes através da maior frequência/média aproximada da média das características comuns. Por fim, a caracterização dos perfis de apólices entre segmentos comuns de forma a compreender como se relacionam os segmentos entre si.

## **b) Importância e Justificação do Problema**

Devido à necessidade ou obrigatoriedade de realização de seguros em várias áreas da sociedade, as seguradoras possuem em carteira elevado leque de produtos abrindo as portas a comportamentos fraudulentos em todas essas áreas, o que tem um custo adicional para todos os restantes clientes considerados cumpridores. Dados estatísticos apontam que no total, os custos com a fraude nas seguradoras atingem o valor anual de 96.8 mil milhões de USD em todo o mundo, onde cerca de 3 mil milhões USD apenas na fraude automóvel. (Erie Insurance, 2012).

A própria existência das seguradoras como um mercado permitiu que o mesmo se tornasse um palco para fraudes e abusos, dado que as seguradoras não reconhecem que o aumento do número de fraudes recai sob o principal problema: o aumento de oportunidades para o crime, aquando da venda de produtos por parte da seguradora, sobre os quais os tomadores de seguros não conseguem conhecer, na sua totalidade, as condições inerentes ao produto pretendido. (Lesch and Brinkmann, 2011; Niemi 1995). As cláusulas contratuais também foram passivas no que diz respeito à não cobertura de

possíveis oportunidades de fraude, principalmente os termos e condições liberais que levaram a que os segurados se tornassem mais relaxados em relação às suas propriedades e bens, para o qual celebraram o contrato. (Niemi 1995).

No entanto, o paradigma alterou e surgiram várias abordagens no sentido de identificar e detectar estes casos *apriori*, numa tentativa de mitigação da revindicação fraudulenta de sinistros ao nível de modelos estatísticos, tecnologias, aplicações de inteligência artificial e *machine learning* levando a diversos problemas na aplicação dos métodos referidos, visto que a detecção dos casos de fraude não é 100% perfeita. (Derrig 2002). Então, questões de carácter ético e moral como: “Serão as apólices classificadas como fraudulentas realmente verdadeiras?” e: “Serão as apólices não identificadas pelo método de detecção realmente não fraudulentas?” moralizaram os investigadores continuar a sua investigação em busca da diminuição dos erros nos seus modelos propostos (Artís et. Al. 2002).



## II. Revisão Literária

### a) Fraude nos seguros

#### a. Contextualização

Um seguro é um contrato civil ou apólice<sup>2</sup>, celebrado entre um indivíduo ou uma entidade<sup>3</sup> e uma companhia de seguros (ou seguradora), no qual esta se responsabiliza a dar cobertura financeira contra um acontecimento de perda, transferindo assim o risco na totalidade para a seguradora. Para tal a seguradora cobra um prémio ao cliente, consoante o risco e a probabilidade de ocorrência associados ao evento de perda, sendo este contrato assinado com base no facto de que tanto a seguradora como o seu tomador de seguro agirão, perante as suas responsabilidades, de boa-fé<sup>4</sup>, isto é os agentes económicos não irão alterar o seu comportamento durante o prazo em que o contrato se encontra em vigor, isto é de acordo com os diferentes contextos económicos, exemplificando: na ocorrência do sinistro segurado, relatar os factos de forma incorrecta de forma a tentar ganhar um benefício financeiro adicional para si ou para terceiros (Lesch and Brinkmann, 2011; Viaene e Dedene, 2004).

A fraude nas seguradoras<sup>5</sup> pode ser considerada um problema moral-psicológico, moral-social e criminal, pois envolve ganhar vantagem da promessa contratual da seguradora de pagar, na totalidade ou parcialmente, em caso de perda. Este problema difere, por exemplo, da evasão fiscal ou do roubo de bens materiais, dado que esta relação contratual única tem implicações importantes para o carácter das apólices fraudulentas, fazendo com que as seguradoras e a sociedade em geral estejam sempre à procura de

---

<sup>2</sup> Documento escrito, datado e assinado formalizando o contrato de seguro. Este inclui as condições sejam estas gerais, especiais e caso houver particulares, acordadas entre a seguradora e o assegurado. (No caso português, conforme a publicação no Decreto-Lei n.º 72/2008 em Diário da Republica)

<sup>3</sup> Visto que, Lesch e Brinkmann (2011), classificam o mundo das seguradoras como um “mercado”, é possível a seguinte denominação.

<sup>4</sup> Segundo a literatura anglo-saxónica, o não cumprimento da boa-fé por parte dos intervenientes dá origem ao termo *moral-hazard*, ou seja o extravio do comportamento ou da boa vontade no seguimento da duração do contrato, sendo possível aumentar as probabilidades do pagamento do valor referenciado na apólice após a ocorrência do sinistro (Lesch and Brinkmann, 2011).

<sup>5</sup> Existem várias denominações para o mesmo problema como fraude contra os seguros (Instituto Seguros de Portugal) ou o termo anglo-saxónico de *insurance fraud*.

novas formas de a combater. Este problema já foi analisado por investigadores de áreas como a economia, seguros, sociologia, criminologia e psicologia, realçando aspectos distintos em comportamentos desonestos. A nível regulamentar, a fraude nas seguradoras está incluída no Pilar I do acordo Solvência II, no cálculo do *Solvency Capital Requirement*, na parcela de Risco Operacional<sup>6</sup>. (Tennyson, 2008; Dionne e Wang, 2013)

A fraude tem origem na violação da apólice assinada entre a seguradora e o cliente, sendo que esta pode ocorrer em qualquer fase da duração da mesma, a partir de acções ou omissões deliberadas, efectuadas pelo segurado, com o intuito de obter das seguradoras ganhos financeiros ou outros, a que não tem direito, para si ou para terceiros, utilizando para isso falsas informações e induzindo assim a seguradora em erro (Brites, 2006; Gilberto, 2010).

O incumprimento do contrato por parte do segurado insere-se na violação do princípio da boa-fé implícito no mesmo, dando lugar a um *moral-hazard*<sup>7</sup>, que se pode traduzir quer numa fraude oportunista<sup>8</sup> – aproveitamento de um evento legítimo executando um *buildup*<sup>9</sup>; quer numa fraude planeada<sup>10</sup> - premeditação de um evento, tornando-o deste modo ilegítimo e, por consequência, originando perdas para a seguradora, uma vez que o custo da fraude leva a inflações nos prémios das apólices de outros clientes e aumenta a exposição ao risco de eventos, devido a restrições contratuais e ineficiências no mercado segurador. O facto de ser possível lucrar perante a fraude leva a sociedade a reflectir sobre as companhias de seguro no entanto, esta tende a acreditar que

---

<sup>6</sup> O risco operacional tornou-se uma importante fonte de risco para as instituições seguradoras, sendo que o mesmo representa entre 5% e 20% do risco total de uma instituição. (Dionne e Wang, 2013)

<sup>7</sup> Problema reconhecido como um risco de negócio intrínseco, onde são tidos em conta tanto a natureza do contrato, a legislação em vigor e as características de activação de apólices da seguradora (Baker, 1996).

<sup>8</sup> Termo utilizado para representar a fraude efectuada por cidadãos comuns, onde não existe um planeamento concreto do evento e sem que, em muitas vezes, os mesmos tenham conhecimento de que estão praticar uma violação contratual, onde sejam causas as mais comuns a “invenção” de perdas ou um pequeno exagero da realidade. (Lesch and Brinkmann, 2011; Tennyson, 2008)

<sup>9</sup> Nomenclatura anglo-saxónica para definir o exagero dos danos causados.

<sup>10</sup> Termo utilizado para caracterizar a fraude cuja qual é principalmente efectuada por grupos associados a actividades criminosas, onde o evento é planeado meticulosamente com o objectivo delineado.

a fraude nas seguradoras é um crime sem vítimas. (Niemi, 1995; Dean, 2004; Miyazaki, 2008, Gilberto, 2010; Tennyson 2008).

A dimensão social da fraude, ou a sua tolerância, depende da atitude por parte dos consumidores contra a mesma, visto que os mesmos possuem o conhecimento do retorno dos seus actos e estes poderão criar estigmas sociais e/ou custos físicos relacionados com a fraude. Tal como no meio social, as atitudes de um indivíduo são baseadas nas atitudes da sua rede de influência, isto é quanto maior as atitudes a favor da fraude, assim como a sua “naturalidade” e tolerância na sociedade, maior será a probabilidade deste adquirir o mesmo comportamento, idealizando assim probabilidades baixas para penalidades e detecção de comportamentos fraudulentos. (Tennyson 2008).

Visto que Portugal é conhecido pela “arte” de enganar a autoridade tributária, a sociedade portuguesa tolera a fraude nas seguradoras, dado que presume que as seguradoras detêm muito capital. Contudo esta ideia é uma falácia, pois as entidades seguradoras pagam milhões de euros em indemnizações, resultantes de ardilosos esquemas, muitas vezes extremamente complexos e altamente bem elaborados. (Martins, 2006).

A fraude nas seguradoras pode ser medida em quatro formas, (1) fraude criminal ou *hard fraud*, (2) suspeita de fraude criminal, (3) *soft fraud* ou abuso sistemático e por fim, (4) suspeita de fraude ou de abuso sistemático. Existe assim, possibilidade de um cliente ser processado por *hard fraud*, sempre que os actos praticados possuam determinadas características como: (1) ser um acto de consciência plena, (2) ser um acto contra a lei<sup>11</sup>, (3) sempre que haja um ganho financeiro, por parte do assegurado e por fim (4) quando exista uma representação material indevida, isto é um “excesso” de declaração do sinistro. Basta uma destas condições não ser cumprida, para se entrar na chamada “zona cinzenta” ou *soft fraud*, onde não existem evidências para um crime e consequência tomada das suas

---

<sup>11</sup> Entenda-se por lei a Norma Regulamentar N.º 10/2009-R, de 25 de Junho elaborada pelo Instituto de Seguros de Portugal onde a fraude nas seguradoras tem como definição: “*Fraude contra os seguros - prática de actos ou omissões intencionais, ainda que sob a forma tentada, com vista à obtenção de vantagem ilícita para si ou para terceiro, no âmbito da celebração ou da execução de contractos de seguro ou da subscrição de operações de capitalização, designadamente os que visem uma cobertura ou pagamento indevido.*”.

acções, mas existe sim um abuso que não justifica a activação dos meios judiciais. (Sparrow, 1996; Derrig et. Al., 2006; Lesch and Brinkmann, 2011).

Lesch e Brinkmann (2011), referem que os segurados têm sido o principal alvo no estudo sobre a fraude nas seguradoras, sendo estas vistas como entidades reguladoras e tribunais. No entanto, as seguradoras embora tenham a capacidade de saber que existem actos de má-fé, falham ao provar a veracidade dos mesmos e outros comportamentos não éticos, a não ser quando activam os mecanismos legais existentes nas transgressões à lei. Exemplo: quando um segurado tem um acidente automóvel e dirige-se a uma oficina para arranjar o veículo, para além de arranjar o que foi estragado no acidente, arranja algo que já se encontrava avariado. Na factura final, vai aparecer que foi arranjado, por exemplo, o pára-choques, com um preço final de 100€, contudo o preço real de arranjo do pára-choques é mais baixo que o apresentado na factura.

Baldock (1997), propôs outro tipo de divisão de clientes fraudulentos, estes podem-se dividir entre (1) o *average offender*, caracterizado pelo cidadão cumpridor da lei, que pertence a qualquer classe social e que, por tentação ou dificuldade financeira, defrauda a seguradora com um tipo de fraude mais leve; (2) o *criminal offender*, o cidadão que, para além de praticar fraude nas seguradoras, comete outro tipo de crimes, continuando o trajecto criminal até ser apanhado e (3) o *organised crime offender*, cidadão que faz do crime uma carreira, proveniente de redes criminosas organizadas e investindo muito tempo e esforço na organização de esquemas fraudulentos complexos, sendo o seu objectivo o extravio de grandes quantidades de dinheiro das companhias de seguro, para financiar outras actividades criminosas.

Segundo Clarke (1989), é possível categorizar os clientes fraudulentos como (1) oportunistas, isto é indivíduos que se aproveitam de uma situação real, sobrevalorizando o valor da participação do sinistro, por exemplo; (2) amadores, cidadãos que efectuem seguros múltiplos e participações inteiramente falsas, mas não em números significativos, e por fim (3) os criminosos profissionais que são indivíduos pertencentes a grandes organizações criminosas, cujo objectivo é praticar fraudes organizadas nas seguradoras, assim como outros tipos de crime.

Com o aumento de número de técnicas e ferramentas de detecção, chega também o aumento do número de casos identificados de fraudes contra as companhias de seguro e consequentemente, o aumento do problema da fraude. Com base neste ponto de vista, Tennyson (2008) afirma que apesar do aumento real dos números da fraude, a sua subida exponencial prende-se com o facto de existir uma maior fiscalização<sup>12</sup>, questionando assim os benefícios no aumento dos casos de fraude, ignorando o elevado custo dos comportamentos fraudulentos para a sociedade contemporânea.

Posto isto, a probabilidade de existência de uma apólice com características fraudulentas não é consensual, existem inúmeros factores socioeconómicos que deverão ter-se em conta nos casos da “zona cinzenta” referenciada por Sparrow (1996), como qual o tipo de segurado que a pratica, quais as fraudes a incluir (ou seja uma fraude oportunista ou planeada), quais são as áreas geográficas com maior incidência, a natureza da perda, o contexto do seguro, a prática empresarial, o género do segurado, o rendimento económico, a condição social, o nível de escolaridade, os impostos como o IRS, IRC, o conhecimento do sistema no geral, os factores psicológicos, a noção de justiça, de confiança e obediência, as percepções negativas das companhias de seguros, assim como os clientes que obtiveram experiencias negativas com as mesmas (Derrig et. Al., 2006; Cialdini, 1989).

As entidades seguradoras possuem ao seu alcance um número vasto de ferramentas analíticas, mas principalmente a matéria-prima da investigação: os dados sobre os seus clientes suspeitos ou não de comportamentos ou atitudes fraudulentas. Sejam estes dados obtidos através da sua relação, quer através dos termos acordados nas cláusulas da apólice, ou mesmo através de relações com entidades externas à seguradora, estes dados facilitam a investigação das seguradoras de todos os casos que achem suspeitos e na tomada de acções internas contra os casos fraudulentos, sem entrar com actos criminais, evitando assim burocracias desnecessárias, pois tanto os recursos públicos (polícia ou tribunais), como os recursos privados de investigação são insuficientes na altura de obter provas concretas, no que diz respeito à investigação a possíveis casos de fraude. (Niemi 1995).

---

<sup>12</sup> Com o aumento da fiscalização atualmente, detetou-se um maior número de casos de fraude. Não existindo fiscalização, estes casos não seriam conhecidos.

Segundo Tennyson (2008), a audição e detecção de casos de fraude diminui a relação de confiança entre a seguradora e os seus clientes, limitando assim os ganhos com as trocas nos seguros e consequentemente aumentando os casos de fraude na própria seguradora, sendo que os consumidores com actos fraudulentos são atraídos para as seguradoras mais brandas com a problemática da fraude. Contudo, dado que a detecção de casos de fraude está relacionada com o *trade-off* entre o efectuar ou não uma auditoria e o seu retorno, seja a nível de recursos (tanto humanos como de capital), seja a nível de custos com processos jurídicos ou até mesmo na facilidade em encontrar a “verdade”, as companhias de seguro revelam sempre um grande interesse em detectar os seus casos de fraude. Contudo, poucos casos são reportados oficialmente, sendo que os benefícios directos da investigação de casos fraudulentos revelam uma diminuição dos custos associados com as mesmas. No que toca a benefícios indirectos existe um maior incentivo a que os clientes diminuam as tentativas de fraude. (Lang e Wambach, 2013; Tennyson 2008).

Niemi (1995) refere que para minimizar o impacto financeiro decorrente da existência de casos fraudulentos, é do interesse da seguradora chegar a um acordo com o cliente fraudulento, para que minimize ou desapareça por completo o impacto financeiro para ambas as partes, seja através das custas judiciais ou mesmo conseguindo que o lesado pague, na totalidade ou parcialmente, a indemnização recebida pelo acto fraudulento, adicionando o facto da companhia de seguros lesada ser capaz de prever futuras acções, por parte de um cliente já identificado como fraudulento. Caso se confirme a existência de fraude, o autor sugere que, para além da rescisão imediata de todas as apólices existentes entre a seguradora e o cliente fraudulento, a seguradora deverá fazer tudo ao seu alcance para pagar o mínimo ao seu segurado até que as provas concretas sejam apresentadas. o pagamento apenas dos danos reais para o qual a apólice original foi efectuada ou o pagamento por parte do segurado da diferença do valor do prémio pago, o não pagamento de qualquer prémio por parte da seguradora, a rescisão do contrato e as provas apresentadas para criminalizar o responsável ou ainda, caso apenas exista fortes suspeitas de fraude, mas não provas concretas.

Em conclusão, a fraude nas seguradoras é um fenómeno complexo e vasto, em virtude das inúmeras características que esta compõe e das acções da mesma e para o mesmo resultante, sendo o mesmo derivante da expansão exponencial tanto da indústria seguradora como da sociedade em geral.

## **b. Enquadramento histórico**

A humanidade sempre sentiu necessidade de encontrar meios que permitissem a tanto a sua protecção, como a protecção dos seus bens, contra a incerteza do futuro. Quando a actividade humana estava ligada a 100% à natureza, totalmente alheia às actividades humanas, as sociedades mais organizadas comprometeram-se a encontrar meios que diminuíssem os riscos associados às suas actividades. Com o aparecimento dos seguros, surgiu também a fraude associada aos mesmos, apesar de apenas nos últimos 30 anos a sua investigação ter sido bastante mais aprofundada.

Existem manuscritos e relatos provenientes da Grécia Antiga<sup>13</sup> que referem o afundamento propositado de embarcações, também conhecido como *ship scuttling*, o qual ficou presente na sociedade ocidental, até ao século XVIII, onde se tornou insustentável para a indústria seguradora. Atingiu níveis de insustentabilidade tão elevados que, no caso da Inglaterra, os indivíduos cujas acções fossem consideradas verídicas, enfrentavam a pena de morte. Tal punição pelo crime originou uma diminuição neste tipo de fraude, levando ainda os consumidores fraudulentos a investir noutro tipo de seguros, nomeadamente em seguros de vida. Já nos séculos XIX e XX, com o crescimento das necessidades da indústria ferroviária e da indústria automóvel houve um crescimento exponencial tanto da indústria seguradora, como dos esquemas fraudulentos associados aos mesmos. Já nos finais do século XX e início do século XXI com a expansão exponencial do mercado segurador a outros sectores como o financeiro, imobiliário, tecnológico, transportador entre outros, novas formas e oportunidades de fraude surgiram desse rápido crescimento (Viaene e Dedene, 2004; Gilberto, 2010; Niemi, 1995; Derrig, 2002)

O interesse pela fraude nas seguradoras também aumentou durante a década de 80 do século XX, em diversos países. Nos Estados Unidos da América foi criada a *special investigative unit* (SIT), no Canadá foram estabelecidas as bases conceptuais para a análise

---

<sup>13</sup> Período datado entre 1100 A.C. e 146 A.C após dominação e inclusão por parte do império romano.

da fraude, no Reino Unido a fraude nos seguros de viagem, automóvel, de propriedade e de negócios tornava-se um problema que necessitava de atenção concertada e planeada levando a que os oito países<sup>14</sup> industrializados subscrevessem um consenso generalizado sobre o problema da fraude, que contudo obteve várias e diversas respostas já no final da década de 80 (Derrig, 2002; Clarke, 1989; Clarke, 1990). Contudo, só em 2006 foi criado em Portugal a CTFRAUDE<sup>15</sup> pela Associação Portuguesa de Seguros, cujo objectivo é a luta contra a fraude nas seguradoras. Hoje em dia, a Guarda Nacional Republicana (GNR), através dos seus núcleos de investigação criminal<sup>16</sup> e investigação de crimes em acidentes de viação<sup>17</sup> permite, em colaboração com a Associação Portuguesa de Seguradoras e as restantes forças de segurança como a Policia de Segurança Publica (PSP) e Policia Judiciaria (PJ), investigar possíveis casos de fraude nas seguradoras em todas as vertentes. O objectivo desta investigação é recolher informação que permita um estudo exaustivo do fenómeno, das suas causas, das suas formas e dos efeitos colaterais, para uma melhor compreensão e acção na prevenção, desencadeando consequentemente uma acção penal célere e delineando estratégias de investigação dirigidas correctamente, investigação essa que se estende a nível internacional<sup>18</sup>, através do gabinete nacional da INTERPOL e da unidade nacional da EUROPOL. (Gilberto, 2010; GNR, 2013).

### **c. Compreensão jurídica**

O termo fraude nas seguradoras está muitas vezes associado ao engano e ao abuso por parte dos segurados face ao seu contrato de boa-fé, celebrado com a entidade seguradora, não constituindo assim um crime. A fraude nas seguradoras é um conceito atribuído internacionalmente (a legislação portuguesa não especifica), seja no campo civil como criminal, o que não opõem a que os actos fraudulentos possam assumir ou não características criminais em dependência dos factos verificados, como por exemplo a realização de outros crimes para levar a cabo o abuso como a falsificação de documentos ou a associação criminosa. (Viaene & Dedene, 2004; Brites 2006; Correia 2010)

---

<sup>14</sup> Sendo estes o Canadá, a França, a Inglaterra, a Irlanda, a Itália, a Alemanha, a Holanda e os Estados Unidos

<sup>15</sup> Acrónimo para Comissão Técnica da Fraude.

<sup>16</sup> Este núcleo tem por competência proceder à investigação de crimes para os quais a GNR tem competência e que não esteja atribuído a outros órgãos.

<sup>17</sup> Competência para investigar crimes resultantes de acidentes de viação que originam feridos graves ou vítimas mortais.

<sup>18</sup> Ao abrigo da lei da organização de investigação criminal – Lei nº49/08, de 27 de Agosto, Artº 12º.



A tipificação autónoma do delito prende-se com a frequência com que a modalidade de burla é verificada, assim como a sua relevância para a sociedade. Visto que em Portugal o crime de fraude nas seguradoras não se encontra enquadrado no código penal português, existe a possibilidade de este se desmultiplicar em diversos artigos do Código Penal, como os crimes de burla simples e qualificada – artigos 217º e 218º, burla relativa a seguros<sup>19</sup> – artigo 219º, o crime de falsificação de documentos – artigo 256º, o crime de falsificação de notação técnica – artigo 258º, o crime de atestado falso – artigo 260º, o crime de associação criminosa – artigo 299º e o crime de simulação de crime – artigo 366º. (Brites 2006; Correia 2007, 2010). No entanto, a problemática da necessidade autónoma do delito em questão mantem-se e a suportar este facto mantem-se o facto do artigo 219 do código penal não suportar na totalidade das ofensas cometidas no âmbito da fraude nos seguros, como por exemplo os casos em que o sujeito activo determina outrem à prática dos actos supra; o facto de não contemplar uma estrutura idêntica ao crime de burla e por fim a contemplação dos crimes de dano e o crime de ofensa à integridade física, alíneas a) e b) do ponto 1. (Costa, 1999).

A nível civil, o Decreto-lei nº72/2008, de 16 de Abril é bastante claro relativamente às punições a tomar caso se verifiquem condutas ou comportamentos fraudulentos por parte do cliente, designadamente “omissões ou inexactidões dolosas” – artigo 25º; “Omissões ou inexactidões negligentes” – artigo 26º; comportamentos de má-fé na omissão de cessação do risco – artigo 44; a realização de actos dolosos, presente nos artigos 46º, 141º, 144º e 148º; a exclusão do segurado e cessação do contrato – artigos 83º e 84º; comportamento doloso assegurado em caso de sinistro ou agravamento do risco – artigo 94º; falta dolosa dos deveres relativos à participação do sinistro – artigos 100º e 101º; incumprimento do dever civil de salvamento – artigo 126º; omissão fraudulenta dos deveres de informação sobre a existência de múltiplos contractos de seguros – artigo. 133º; ato ou omissão que prejudique o direito de sub-rogação do segurador – artigo 136º; a pluralidade de lesados –

---

<sup>19</sup> O artigo refere os seguintes pontos: 1 - Quem receber ou fizer com que outra pessoa receba valor total ou parcialmente seguro: a) Provocando ou agravando sensivelmente resultado causado por acidente cujo risco estava coberto; b) Causando, a si próprio ou a outra pessoa, lesão da integridade física ou agravando as consequências de lesão da integridade física provocada por acidente cujo risco esteja coberto; é punido com pena de prisão até 3 anos ou com pena de multa. 2 - A tentativa é punível. 3 - O procedimento criminal depende de queixa. 4 - Se o prejuízo patrimonial provocado for: a) De valor elevado, o agente é punido com pena de prisão até 5 anos ou com pena de multa até 600 dias; b) De valor consideravelmente elevado, o agente é punido cm pena de prisão de 2 a 8 anos. 5 - É correspondentemente aplicável o disposto no artigo 206º (Restituição ou reparação)

artigo 142º e 180º e por fim situações de suicídio, homicídio e danos corporais provocados, artigos 191º, 192º e 193º respectivamente.

É possível observar também os efeitos que surgem sob a forma de nulidade/anulabilidade do contrato, a recusa de pagamento e/ou a redução da prestação concedida, o exercício do direito de regresso ou a negação da cobertura por parte da instituição seguradora. É relevante destacar ainda a instituição da Norma Regulamentar nº 10/2009-R, em 25 de Junho de 2009 (ISP), ao abrigo do artigo 131º-F do Decreto-Lei 94-B/98, de 17 de Abril, aditado pelo Decreto-Lei 2/2009, de 5 de Janeiro, estabelecendo a obrigatoriedade das empresas seguradoras implementarem práticas antifraude, sendo estas baseadas na detecção, prevenção e denuncia de situações de fraude nas seguradoras, assim como uma definição regulamentar da mesma, estando a sua génese aplicada a qualquer fase do ciclo de vida da apólice e não somente à ocorrência do sinistro, como acontece actualmente. (Brites, 2006; Decreto-Lei n.º 72/2008, de 16 de Abril)

Em síntese, é possível analisar a complexidade judicial do fenómeno da fraude nas seguradoras em Portugal e verificar existe ainda muito trabalho a fazer no que toca à regulação deste crime por parte das entidades jurídicas.

## **b) A fraude no seguro automóvel**

O seguro automóvel é um dos produtos das seguradoras, no ramo não vida, mais afectado pela utilização de práticas fraudulentas. Existem vários factores que contribuem para esta situação, primeiro pelo facto de ser um seguro de cobertura obrigatória<sup>20</sup>, atendendo às características do veículo, do tipo de contrato, do produto adquirido com multiplicidade de coberturas e à necessidade de cobertura de danos materiais e corporais causados a terceiros e a ocupantes do veículo (com excepção do condutor) e por último pelo facto de estar presente num mercado completamente massificado em toda a sociedade, que é o sector automóvel. (Derrig, 2002; Viaene e Dedene, 2004; ISP, 2011).

Estudos efectuados por Tennyson (1997 e 2002) demonstram que o seguro automóvel é dos produtos mais vendidos e consequentemente participados, classificado pelos segurados como sendo o seguro onde a probabilidade de fraude é a mais elevada.

---

<sup>20</sup> Em Portugal, tornou-se obrigatório a partir de 1 de Janeiro de 1980, o seguro de responsabilidade civil automóvel, publicado no decreto-lei nº 408/1979 de 25 de Setembro, onde no presente dia se encontra no enquadramento legal no decreto-lei nº 291/2007, de 21 de Agosto.

Em Portugal não existem quaisquer indicadores precisos sobre a fraude nas seguradoras, nomeadamente quanto à sua extensão, assim como estudos publicados sobre este assunto, visto que só durante a primeira década do século XXI é que as empresas de seguros despertaram para um dos verdadeiros problemas da actividade seguradora, que é a fraude nos seguros automóveis. (Martins, 2006).

Assim como a fraude nas seguradoras em geral, a fraude nos seguros automóveis pode dividir-se em (1) fraude oportunista (*soft fraud*) e (2) fraude planeada. Na fraude oportunista predominam as acções não verídicas, como a omissão de factos relevantes com o intuito de induzir a entidade seguradora em erro e com o objectivo de ganhar financeiramente de forma indevida, em qualquer fase do contrato, assim como a fraude planeada, (*hard fraud*). É possível caracterizar a fraude oportunista em todas as fases da duração do contrato, seja na fase de subscrição da apólice, onde são efectuadas declarações de informações falsas, na assinatura do contrato, destacando dados pessoais incorrectos seja a data de nascimento, o historial do condutor ou a residência do mesmo e também na fase de subscrição da apólice após a ocorrência do sinistro, para que o mesmo seja participado logo de seguida. Isto é, durante a fase de informação/participação do sinistro surge a sobrevalorização no que diz respeito ao valor dos danos corporais e/ou materiais, a participação de danos anteriores ao sinistro, a falsificação de facturas e a deturpação dos factos ocorridos. No que toca à fraude planeada, é possível verificar a participação de sinistros que nada têm a ver com as coberturas incluídas no contrato de seguro. São participados casos como roubo, incêndio ou furto, a participação de um sinistro ocorrido que fora efectuado premonitoriamente e por fim, a troca do número de identificação do veículo (NIV), que se baseia na venda de um veículo acidentado, sendo este reportado à seguradora como reparado e, de seguida a utilização do NIV deste num carro roubado, da mesma marca e modelo (Tennyson, 2002; Brites, 2006; Derrig et. Al., 2006; Gilberto, 2010; Lesch and Brinkmann, 2011).

É importante salientar que para perpetuação de acções fraudulentas contribuem também factores como a cumplicidade entre o segurado e as oficinas de automóveis (aquando da existência de danos materiais); e/ou a cumplicidade entre o segurado e os serviços médicos (no caso dos danos corporais) que leva à sobrevalorização dos danos reportados no sinistro, também os danos corporais do foro psicológico levam a uma

sobrevalorização, visto a existência da dificuldade da avaliação na existência do dano participado; o número de sinistros contemplando apenas uma viatura (elevando assim a probabilidade de não haver qualquer testemunha) e por fim o número elevado de carros roubados aquando da cobertura do mesmo pelo seguro (Clark, 1989; IRC, 1991 e 1993; Cummins e Tennyson, 1996; Carroll e Abrahamse, 2001; Dionne e Gagné, 2002; Tennyson, 2008; Gilberto, 2010).

Existem assim várias técnicas analíticas (Figura 1), para a detecção de casos de fraude nas seguradoras, enumerando (i) as regras de negócio, onde se inclui a estratificação dos números para detectar dados não usuais, bem como a junção de dados de diferentes fontes para identificar padrões; o teste de valores duplicados e a validação de datas, para a identificação de datas incomuns, (ii) a detecção de anomalias, onde vários indicadores de performance chave são elaborados com determinados limites, onde os alertas tocam quando estes são ultrapassados, por exemplo a detecção de *outliers*, (iii) a utilização técnicas de *data mining*, como o *clustering*, a classificação, a previsão, a regressão ou a visualização (iv) a utilização de *text mining* e por fim (v) a implementação de *Social Network Analysis* ou redes, ou seja a implantação de redes ou relações entre indivíduos e entidades suspeitas de fraude. (Balasubramanian & Sithic, 2013; ACL, 2014; SAS, 2012).

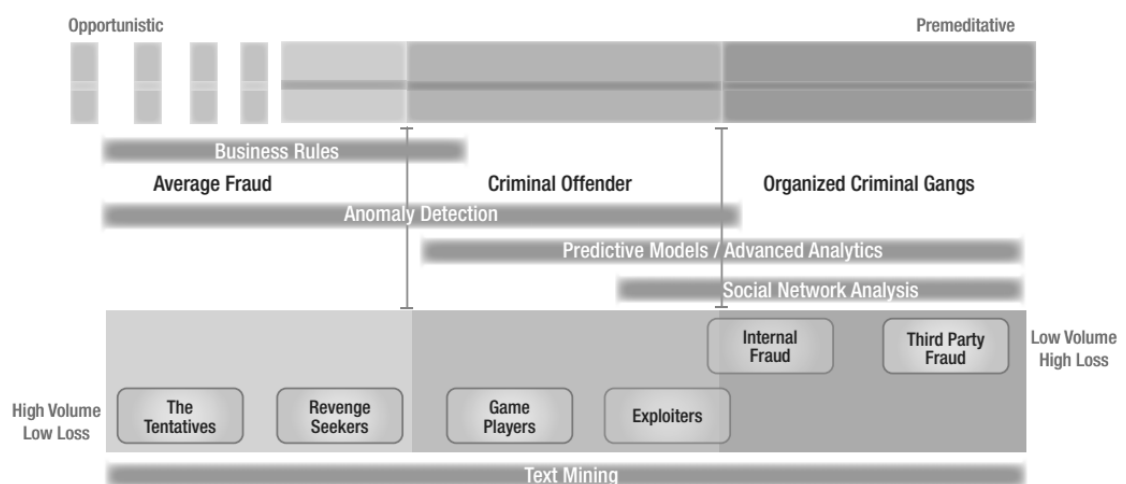


Figura 1 – Aplicação de técnicas antifraude para os diferentes casos de fraude nas seguradoras. (SAS, 2012)

Em resumo, a complexidade do fenómeno da fraude no seguro automóvel acarreta diversos problemas associados, os quais demonstram que a sociedade, apesar das recentes evoluções no campo jurídico, não consegue lidar com este problema da forma mais eficaz e eficiente.

## **c) Redes**

### **a. As ciências da complexidade**

A criação de redes entre indivíduos deriva das ciências da complexidade ou seja, sistemas naturais que apresentam um comportamento estrutural, onde não é possível prever facilmente a dinâmicas dos agentes que os compõem de forma isolada. Apesar de estas possuírem características estruturais comuns, exemplificando algumas ideias unificadoras como a teoria limiar do caos ou a criticalidade auto-organizada, não é de todo possível obter uma teoria unificada satisfatória que descreva os sistemas complexos e toda a sua complexidade, pois existem pelo menos duas propriedades consensuais em relação à sua caracterização, sejam estas a dificuldade de previsão do comportamento do sistema e a criação da estrutura de relações. Neste contexto, são considerados sistemas complexos uma vez que os sistemas simples<sup>21</sup> apresentam resultados erráticos e imprevisíveis e também os quais apresentam uma complexidade associada à criação de estruturas derivada da caracterização do comportamento emergente dos sistemas complexos<sup>22</sup>, pois as estruturas criadas são tão ricas e diversas quando o sistema se encontra num processo de transição para a imprevisibilidade absoluta, onde é possível caracterizar a ocorrência de transições de fase a quando existem pequenas variações nos parâmetros quantitativos do sistema leva a grandes variações qualitativas, o que também pode ser descrito como a lei de potencia (Mendes, 1998)

Os sistemas complexos são orientados por várias propriedades fundamentais, entre elas a interdependência<sup>23</sup> dos elementos do sistema, a emergência<sup>24</sup> de novas propriedades

---

<sup>21</sup> São considerados sistemas simples os quais apresentam poucos graus de liberdade e são determinados por equações bem delineadas.

<sup>22</sup> São considerados sistemas complexos devido ao seu elevado número de graus de liberdade.

<sup>23</sup> A reacção de cada um dos agentes está fortemente condicionada com o comportamento dos restantes, sendo que é muito difícil de prever.

<sup>24</sup> Por exemplo, a auto-organização dinâmica, que é um termo utilizado para descrever a relação da dinâmica colectiva com as dinâmicas individuais.

do sistema como um todo, a não-linearidade<sup>25</sup>, a caoticidade<sup>26</sup>, a universalidade<sup>27</sup> e por fim a auto-semelhança, que correspondem ao resultado de um processo recursivo onde as interacções entre os elementos do sistema complexo criam novas estruturas também complexas. Apesar de este conceito ser um dos mais populares na área dos sistemas complexos, devido principalmente à sua participação na descrição de objectos fractais, é a auto-dessemelhança que melhor captura a complexidade natural dos sistemas, devido à informação adicional necessária para quantificar o sistema numa nova escala. (Araújo, 2006)

Assim, é possível verificar que a problemática das redes surge da necessidade da interpretação de vários fenómenos naturais e humanos, os quais as soluções tradicionais mais simplificadas não conseguem encontrar resposta.

### **b. Contextualização histórica**

A origem do estudo das redes, ou Teoria dos Grafos, remota ao século XVIII onde é iniciada por Leonard Euler e o seu problema das sete pontes de Königsberg<sup>28</sup>. Neste problema existem duas grandes ilhas, que juntas formavam um conjunto de quatro grandes bocados de terras, os quais eram ligados por sete pontes (Figura 2.)

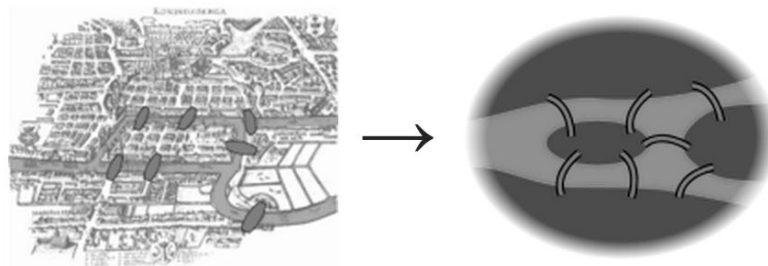


Figura 2 - As sete pontes de Königsberg e a origem da Teoria dos Grafos (Mendes, 2006)

<sup>25</sup> Tem consequências práticas de extrema relevância, principalmente ao nível social e económico, pois a duplicação de uma variável, num sistema que se espera um determinado comportamento, não significa duplicar o valor da resposta. Esta é a propriedade mais utilizada na caracterização de sistemas complexos pois significa que os mesmos não podem ser vistos como uma sobreposição dos comportamentos elementares dos seus constituintes.

<sup>26</sup> Caso particular da criação de estrutura.

<sup>27</sup> Também pode ser descrita através de lei de escala.

<sup>28</sup> Actualmente a cidade chama-se Kaliningrado.

Segundo Euler, não é possível atravessar as sete pontes, apenas uma vez e terminar o percurso na margem escolhida para a partida, onde tal resultado apenas seria possível caso cada margem possuía um número par de pontes. Isto pretende demonstrar assim a diferença entre as abordagens geométrica e topológica<sup>29</sup>, onde é possível verificar que este problema se restringe a uma rede com quatro nodos e sete ligações. (Figura 3.) (Newman, 2003)

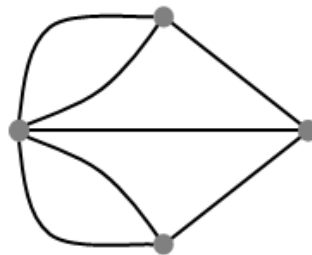


Figura 3 - Ilustração da rede (Mendes, 2006)

Mais recentemente, no século XX, dois matemáticos húngaros, Paul Erdős e Alfred Rényi, iniciam o estudo dos mecanismos de formação de redes aleatórias, concluindo que existem duas classes de redes perfeitamente distintas, as redes regulares e as redes aleatórias. Nas redes aleatórias não existe qualquer padrão no estabelecimento de ligações e as mesmas dependem de dois factores, sendo estes o coeficiente de agregação<sup>30</sup>, ou *clustering* e o caminho médio mais curto<sup>31</sup>, também conhecido como *characteristic path length*. (Araújo, 2006)

Em 1967, o cientista Stanley Milgram concebeu a primeira teoria sobre o que seria conhecido como as redes *small world*<sup>32</sup> (SW), onde o mesmo afirma que todos os indivíduos americanos estavam separados, em média, por apenas seis intermediários. Contudo, a mesma teoria aplicada aos actores de Hollywood, sendo a definição “tendo participado no mesmo filme”, era uma média de apenas três intermediários, e desta forma demonstrou o poder desta nova classe de redes no domínio dos problemas das ciências sociais. (Watts, 1991; Albert e Barabási, 2002; Blass, 2002)

<sup>29</sup> Não incide sobre a forma ou posição exacta dos objectos mas sim sobre as ligações estabelecidas entre os seus elementos.

<sup>30</sup> Probabilidade de, para cada elemento da rede, os pares de elementos a ele ligados estarem conectados entre si.

<sup>31</sup> Número médio de ligações necessárias para ligar cada par de elemento da rede.

<sup>32</sup> Classe de redes onde as mesmas são caracterizadas por reunir propriedades opostas, que poderiam ser consideradas até impossíveis de representar na mesma estrutura, obtém um elevado valor de coeficiente de

Posto isto, é possível verificar que as teorias matemáticas sobre as redes são relativamente recentes na história da matemática, com cerca de apenas 300 anos, e que a sua evolução ficou sempre marcada com a aproximação aos problemas das ciências sociais, tentando desta maneira solucioná-los.

#### **d) Tipologias de redes**

As redes podem ser definidas como a arquitectura da complexidade (Barabási, 2001), e a sua complexidade é arquitectada sobre hierarquias, sendo um tipo específico de rede e cuja ubiquidade é notória, sendo frequentemente encontrada tanto nas ciências sociais como na física ou na economia (Simon, 1969). Para isto, Araújo (2006) contribui com o facto de as redes estarem assentes em duas propriedades principais dos sistemas complexos, a interdependência e a emergência, o que faz com que a análise recaia principalmente sobre as ligações entre os indivíduos, relegando para segundo plano os demais atributos acessórios e complementares.

Ainda segundo Araújo (2006), existem três formas distintas de estudar os sistemas em rede, (1) o primeiro tipo de estudo assenta sobre a estrutura do sistema em rede, procurando quantificar os coeficientes topológicos separadamente, como as redes aleatórias; (2) o segundo tipo de estudo assenta sobre a dinâmica onde se pretende caracterizar quais os tipos de comportamentos mais prováveis a serem observados naquela estrutura, exemplo disso são as redes *small world*, e por fim (3) o estudo da evolução da própria rede que muda com o tempo, como as estruturas *free-scale*.

Mendes (2006) afirma que a criação de estruturas modulares derivam de três mecanismos complementares do processo selectivo, em primeiro lugar o facto de a dinâmica das interacções entre os agentes e entre si e o ambiente origina um número possível de estruturas; em segundo lugar a certeza de que a competição entre as estruturas selecciona a mais viável, com base nas funcionalidades exibidas por certos módulos cooperantes entre si e por fim o caso de, em igualdade de circunstâncias, os simples acasos ou ambientes exteriores, podem originar acidentes que condicionam a permanência ou não das estruturas criadas.

---

agregação devido ao seu baixo valor do seu caminho médio. O interesse por estas aumenta devido principalmente à sua ubiquidade.



Em termos genéricos, uma rede corresponde a um conjunto  $G$ , sendo esta constituída por um par de conjuntos  $G = \{N, L\}$ , onde  $N$  é o conjunto dos nodos da rede,  $N_1, N_2, \dots, N_M$  e  $L$  o conjunto das ligações entre os nodos, sendo que o grau ( $g_i$ ), do nodo  $N_i$ , corresponde ao número de ligações nas quais  $N_i$  participa, o grau ( $g$ ) de uma rede  $G$  corresponde à média aritmética dos graus de todos os nodos de  $G$  e os valores de agregação ( $C_G$ ) e do caminho mais curto ( $P_G$ ) de uma rede  $G$  correspondem efectivamente às médias aritméticas dos valores obtidos para cada coeficiente medido para cada nodo de  $G$ . O valor de  $P_i$ , que é conhecido como o sendo o caminho médio mais curto do nodo  $i$  a todos os outros nodos, é também dado pela média aritmética do numero de conexões necessárias para ligar o nodo em causa aos restantes nodos da rede, sendo que o seu limite inferior é  $P_i \geq 1$  e quando a rede não possui conexões,  $P_i = \infty$ , visto que todos os nodos estão “desligados”. Assim, designando o número de vizinhos do nodo  $i$  por  $v_i$ , correspondendo ao conjunto de nodos ligados directamente ao nodo  $i$ , o coeficiente de agregação  $C_i$  é dado pela Formula 1, Onde  $E(v_i)$  corresponde ao número limite de conexões existentes entre os elementos de  $v_i$ . É de relevância referir ainda que, nas redes aleatórias, os coeficientes  $C_G$  e  $P_G$ , possuem em valor baixo a quando da comparação com as redes regulares, pois as mesmas apresentam um alto coeficiente de agregação, assim como caminho médio. Assim sendo, a existência de um caminho médio curto entre todos os elementos da rede beneficia todas as funcionalidades associadas ao encurtamento de distâncias, sejam elas geográficas ou topológicas, em contraponto, o elevado valor para o coeficiente de agregação revela a presença de ligações redundantes, favorecendo a possibilidade de substituição de funcionalidades no caso de erros ou a avarias referentes a falhas de ligações<sup>33</sup>. (Araújo, 2006)

$$C_i = \frac{E(v_i)}{v_i(1 - v_i)/2} \text{ onde: } 0 \leq C_i \leq 1$$

Formula 1: Coeficiente de agregação genérico

No entanto, segundo Watts (1999), as redes do tipo *small world* são resultantes de um mecanismo que faz a interpolação entre a ordem, que é representada por uma rede absolutamente regular, e a desordem, também considerada como a rede aleatória. Este

<sup>33</sup> Referencia a redes de electricidade ou de transporte domestico.

mecanismo de interpolação tem um único parâmetro  $\beta$ , que significa a probabilidade de cada nodo da rede alterar de vizinhos durante o processo de *rewiring*.

Na Figura 4 é possível verificar que no primeiro anel, onde  $\beta = 0$ , a rede apresenta uma estrutura regular e um coeficiente de agregação alto como nas redes regulares, onde cada nodo está ligado aos seus quatro vizinhos. Para  $0 < \beta < 1$ , é possível visualizar que os nodos vão-se desligando dos seu vizinhos e ligando a outros aleatoriamente assim como uma diminuição progressiva do coeficiente de agregação. Por fim para  $\beta = 1$  verifica-se uma rede onde as suas ligações foram estabelecidas de forma totalmente aleatória, apresentando um coeficiente de agregação tão baixo quanto as redes aleatórias.

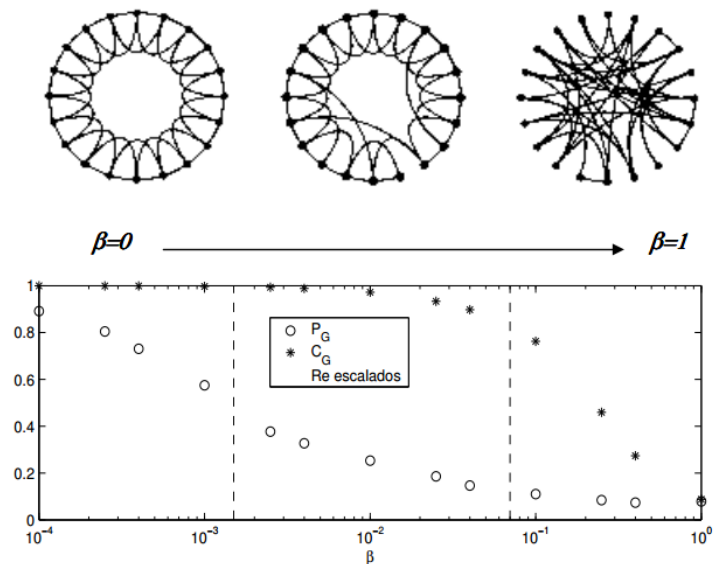


Figura 4 – Descrição visual do processo de *rewiring* (Fonte: Araújo, 2006)

Em termos comparativos, segundo Araújo (2006), tanto nas redes aleatórias como nas redes *small world*, a distância média mais curta entre os nodos da rede é o valor aproximado do logaritmo natural do número de nodos ( $M$ ) da rede (Formula 2), contudo os dois tipos de redes divergem em relação ao coeficiente de agregação, enquanto nas redes aleatórias a probabilidade de dois nodos se encontrarem conectados é dado pelo grau médio da rede (Formula 3).

$$P_{\text{rand}} = P_{\text{sw}} \approx \frac{\ln(M)}{\ln(g)}$$

Formula 2 – Distancia média mais curta.

$$C_{\text{rand}} = \frac{g}{M}$$

Formula 3 - Grau médio da rede.

Todavia, nas redes *small world* o coeficiente de agregação é elevado como nas redes regulares, sendo obtido através de:

$$C_{\text{sw}} = \frac{3}{4} \left( \frac{g-2}{g-1} \right) \text{ onde: } g \geq 2$$

Formula 4 – Coeficiente de agregação das redes *small world*.

No entanto, também as redes com invariância de escala, ou *scale-free*, são caracterizadas por terem um caminho médio curto mais baixo. Estas funcionam através de um mecanismo de crescimento da rede, que é efectuado através do estabelecimento de ligações preferenciais, que de acordo com o mesmo as ligações vão sendo criadas numa rede preferencialmente entres os nodos com um número elevado de ligações, sendo que os nodos vão sendo adicionados à medida que a rede cresce pois a mesma é iniciada com poucos nodos assim como ligações ( $m_0$ ). A cada iteração são adicionados novos nodos e com este novas  $m \leq m_0$  ligações, as quais são estabelecidas entre os nodos recém-criados com os que possuem maior número de ligações, onde o coeficiente  $R(g_i)$ , (Formula 5) significa a probabilidade do nodo  $i$ , de grau  $g_i$ , ser destino ou origem de uma nova ligação. Porem, apesar de se conhecer o expoente característico da distribuição dos graus da rede, para este mecanismo de crescimento com ligações preferências, o mesmo não depende do único parâmetro ( $m$ ) que é empregue na construção da mesma. (Albert e Barabási, 2002)

$$R(g_i) = \frac{g_i}{\sum_j g_j}$$

Formula 5 - Coeficiente  $R(g_i)$ .

Resumindo, verifica-se assim que a criação de redes não é o simples resultado de um mecanismo de selecção sobre um número de transformações aleatórias, mas sim o resultado de um produto complexo onde os parâmetros selectivos, o factor da

aleatoriedade, assim como a auto organização se unem para formar o complexo estrutural em causa.

### III. Metodologia

#### a) Descrição da Investigação

A escolha do método científico que descreve a escolha de acções sistemáticas sobre a qual a situação em estudo seja descrita e explicada assenta em dois factores: (1) a natureza do objectivo da sua aplicação e se o objectivo tem em conta o estudo em questão. (Fachin, 2001) Posto isto, o método científico mais apropriado para a investigação em curso é o caso de estudo pois, o mesmo consiste na observação de um fenómeno que ocorre no seu ambiente natural onde os dados obtidos a partir da sua observação, sendo que para Yin (1994) é aplicável o método do caso de estudo quando o investigador encontra dificuldades em salientar as variáveis que são consideradas importantes para a investigação e quando o objectivo do mesmo é a análise e descrição do fenómeno de uma forma profunda e global.

Então, segundo Yin (1994), um caso de estudo pode ser definido com base nas características do objecto de estudo, assim como as características associadas ao processo de recolha de dados e às possibilidades de estratégias de análise dos mesmos, onde este possui o objectivo de explicação, descrição ou exploração de um fenómeno. O mesmo refere ainda que o plano de investigação do caso de estudo é muito demorado e que o mesmo apresenta falta de rigor apesar de ser possível provar a validade do estudo. Já Fidel (1992) por seu lado, afirma que um caso de estudo pode ser considerado como um método específico da pesquisa de campo, sendo que este tipo de pesquisas são investigações de ocorrências que vão acontecendo independentemente do investigador. Ainda segundo este, o caso de estudo objectiva na compreensão do objecto em análise assim como no desenvolvimento de teorias genéricas sobre o fenómeno estudado.

O objecto de estudo desta investigação centra-se, como referido anteriormente, numa amostra de uma base de dados com registos sobre casos de accionamento de apólices fraudulentas e não fraudulentas de três companhias de seguro do ramo não vida que operam no mercado português, cujos nomes não serão revelados por questões de confidencialidade. A amostra foi obtida através da aplicação de técnicas com base nas regras de negócio e detecção de *outliers*, como demonstrado pela Figura1, sendo esta composta por 355.483 observações com um total de 407 variáveis e extraída entre 2 de

Janeiro de 2009 e 10 de Dezembro 2012. Foram escolhidas para análise as seguintes variáveis da amostra:

<b>Variável:</b>	<b>Descrição Funcional:</b>
<b>NUM_Antig_Policy</b>	Antiguidade da Apólice Original em meses
<b>NUM_Antig_Alt_Policy</b>	Antiguidade da última alteração nas coberturas da apólice
<b>Perc_Policy_Exposure</b>	Valor Seguro total nesta Apólice (Em % do valor mais alto - 25%, 50% e 75%)
<b>VAL_Prop_Exposure</b>	Valor Seguro do Veículo
<b>COD_Claim_Time</b>	Hora do Sinistro M:07h.12h, T:12h.20h, N:00h.07h, PN:20h.24h, S (sem info)
<b>FLG_Claim_Wkend</b>	Se o Sinistro ocorreu ou não durante um Final de Semana ou Feriado
<b>COD_Claim_Loc_Type</b>	Tipo de Local onde se deu o Acidente
<b>COD_Claim_Loc_Nuts</b>	NUT onde se deu o Acidente
<b>FLG_Claim_Overseas</b>	Se o Sinistro ocorreu no Estrangeiro ou em Portugal
<b>FLG_Claim_Insured</b>	Se o Tomador é reclamante
<b>FLG_Claim_Police_Rep</b>	Se existiu ou não Intervenção das Autoridades
<b>FLG_Claim_Injured</b>	Se existiu ou não Danos Corporais
<b>FLG_Claim_Witness</b>	Se existiram ou não testemunhas
<b>COD_NAT_CI</b>	Nacionalidade do Segurado
<b>COD_Gender_CI</b>	Sexo do Segurado
<b>COD_Occupation_CI</b>	Profissão do Segurado
<b>COD_INJ_Nature_CI</b>	Natureza da Lesão no Segurado
<b>COD_VI_MM</b>	Marca e Modelo Veículo Segurado
<b>COD_VI_TYPE</b>	Tipo do Veículo Segurado
<b>VAL_VI_INSURED_VALUE</b>	Valor Seguro do Veículo Segurado
<b>ID_Vehicle_V1</b>	Identificador do Veículo 1
<b>NUM_VI_ENGINE</b>	Cilindrada do Veículo Seguro
<b>NUM_V1_KM</b>	Quilometragem do Veículo 1
<b>NUM_V1_Age_FR</b>	Número de Anos de Antiguidade do Veículo 1 ao Ano de Construção
<b>COD_V1_MM</b>	Marca e Modelo Veículo 1
<b>COD_V1_TYPE</b>	Tipo do Veículo 1
<b>ID_Vehicle_V2</b>	Identificador do Veículo 2
<b>NUM_V2_KM</b>	Quilometragem do Veículo 2
<b>COD_V2_MM</b>	Marca e Modelo Veículo 2
<b>COD_V2_TYPE</b>	Tipo do Veículo 2
<b>NUM_V2_ENGINE</b>	Cilindrada do Veículo 2
<b>NUM_V2_Prior_Lines</b>	Número de Sinistros anteriores envolvendo o Veículo 2
<b>NUM_Days_Incept</b>	Dias decorridos entre o Sinistro e a notificação da ocorrência
<b>FLG_PC_Providers_CI</b>	Indicador se há Sinistros anteriores envolvendo o Segurado e Oficina
<b>FLG_VI_Chose_Garage</b>	Indicador Cliente Veículo Segurado escolheu a oficina
<b>FLG_VI_Repair_PFraud</b>	Indicador se oficina associada ao Veículo Segurado teve fraudes anteriores
<b>FLG_PC_V1_Provider</b>	Indicador se há Sinistros anteriores envolvendo a oficina do Veículo 1
<b>FLG_Susp_Fraud</b>	Indicador de suspeita de fraude
<b>FLG_Conf_Fraud</b>	Indicador de fraude provada

Tabela 1 – Enumeração e descrição das 39 variáveis em utilização

As 39 variáveis, Tabela 1, foram escolhidas segundo a bibliografia, para Derrig et. Al. (2006) e Cialdini (1989) onde os factores socioeconómicos como a profissão do segurado características do automóvel entram em consideração assim como a localização geográfica, local onde ocorre o acidente assim como a fase do dia em que o mesmo acontece, o género do segurado, a natureza da perda e o contexto do seguro que foram também seleccionadas. Também as oficinas e o tipo de seguro surgem aqui como um factor importante no processo de fraude segundo Dionne e Gagné (2002) assim com Tennyson (2008);

A realização deste estudo será efectuada com base na análise e segmentação dos diversos tipos de cliente em quatro segmentos distintos<sup>34</sup> e posterior construção de rede. Esta será construída através da ligação entre elementos compartilhados dos segmentos para assim ser possível analisar as relações semelhantes entre os conjuntos anteriormente criados para a construção mais exacta de um padrão com as características de cliente fraudulento, colocando assim em hipótese as técnicas utilizadas para a detecção pelas instituições seguradoras para verificação de casos de fraude.

---

<sup>34</sup> Os quais o cliente é não-suspeito e não-confirmado, não-suspeito e confirmado, suspeito e não confirmado e suspeito e confirmado.

## IV. Resultados

Como referido no capítulo anterior, esta investigação tem como ponto de partida a análise de uma amostra de uma base de dados proveniente de três seguradoras que opera no mercado de seguros não vida português, identificada a partir de técnicas aplicadas como as regras de negócio e detecção de *outliers*, dando origem a duas variáveis alvo relevantes, o indicador de suspeita de fraude (***FLG\_Susp\_Fraud***) e o Indicador de fraude provada (***FLG\_Conf\_Fraud***).

### a) Segmentação e análise da amostra

O primeiro passo para uma análise mais eficaz passou por segmentar a amostra em quatro subamostras utilizando as variáveis em cima referidas enumerando: (1) os casos onde as apólices não são suspeitas de fraude nem efectuaram efectivamente fraude (***FLG\_Susp\_Fraud*** = 0 & ***FLG\_Conf\_Fraud*** = 0), (2) os casos onde as apólices são suspeitas mas não confirmam fraude (***FLG\_Susp\_Fraud*** = 0 & ***FLG\_Conf\_Fraud*** = 1), (3) os casos das apólices que não são suspeitas mas confirmaram acções fraudulentas (***FLG\_Susp\_Fraud*** = 1 & ***FLG\_Conf\_Fraud*** = 0) e por fim, (4) os casos onde as apólices são suspeitas e efectivamente ocorreram acções fraudulentas (***FLG\_Susp\_Fraud*** = 1 & ***FLG\_Conf\_Fraud*** = 1). Das 355.483 observações iniciais foram distribuídas 353.816 para o segmento 1, 1.539 no segmento 2, 47 presentes no segmento 3 e 81 no segmento 4.

O segmento 1 é caracterizado por possuir uma amostra onde as apólices possuem uma duração média de 2 anos até à ocorrência do sinistro em questão, com as apólices a situarem-se junto do valor médio seguro para o total da base de dados. Este é composto essencialmente por veículos ligeiros de passageiros, 97.4%, pertencerem a classes de gama média como a Ford, a Opel, a Renault e a Peugeot, com um valor médio no intervalo entre 8500€ e os 24500€, uma idade média de 12 anos, uma cilindrada entre no intervalo entre os 750 e 1600 CC e onde o tomador de seguro não é reclamante da apólice em cerca de 97% dos casos. Os sinistros ocorreram maioritariamente durante o período da tarde, entre as 12h e 20h, cerca de 23,5% do total, e durante a tarde com uma percentagem de 19%, ocorrendo em dias de semana, com cerca de 78% do total, em Portugal continental mais especificamente nas zonas do grande Porto e Lisboa, constituindo o seu conjunto 47% do total. Existem também muitos poucos danos corporais, em aproximadamente 3,5% das



apólices reportadas, assim como qualquer intervenção das autoridades, com apenas 0.78% dos reportes obtiveram, ou quaisquer testemunhas do sinistro, apenas 0.1% dos casos tiveram testemunhas, e uma média de reporte do mesmo à seguradora de 24 dias. Em relação às oficinas, todos os segurados escolheram qual a oficina para a qual decidiram arranjar o veículo, que são não reincidentes na escolha, apenas 3.6% dos sinistrados voltaram a escolher oficinas escolhidas em sinistros anteriores, mas onde as mesmas possuem um indicador de como foi comprovado um acto de fraude anterior de cerca de 2.1%. Por fim, os sinistrados caracterizam-se por serem maioritariamente do género masculino, 53.3%, de nacionalidade portuguesa, 73% contra os 0.5% de estrangeiros, com pouca informação sobre a sua profissão, Apenas é conhecida para 2.9% dos casos. As lesões que estes contraíram do sinistro foram maioritariamente lesões superficiais, 34%, contudo referir ainda um elevado conjunto de lesões não especificadas, 46.9%. (Anexos: Tabelas 2 – 26, Gráficos: 1 - 20)

Por sua vez, o segmento 2 caracteriza-se por obter uma duração média por apólice de cerca de 18 meses até à ocorrência e posicionando-se também junto do valor médio do valor seguro pela apólice. Este é composto essencialmente por veículos ligeiros, 94.1%, mas com uma pequena percentagem significativa para os veículos de mercadorias, cerca de 4,5% do total, estes pertencem a uma maioria de gama média alta como a Mercedes e a AUDI, mas também com veículos de gama média como a Opel e a Renault, com valores de mercado entre os 10000€ e os 50000€ na sua maioria, com uma idade média de 6 anos, mas sem informação sobre a sua cilindrada. Os sinistros em questão ocorreram na sua maioria no período compreendido entre as 00h e as 07h da manhã, 43.08% do total contra os 20% que não tem qualquer tipo de informação sobre o horário a que ocorreram e os 1.9% dos que ocorreram no período da tarde entre as 12h e as 20h. Estes surgiram durante os dias de semana, 74%, onde cerca de 50% ocorreram em Portugal continental, mais precisamente nas regiões da grande Lisboa, grande Porto, Cávado, Tâmega, Ave, Península de Setúbal e Algarve, onde em 11% dos mesmos existiram danos corporais mas em todos os casos sem intervenção das autoridades ou quaisquer testemunhas, verifica-se também que sinistrado não é o reclamante da apólice em 100% dos casos. O tempo médio de reporte à companhia de seguros do sinistro foi cerca de 14 dias após a ocorrência do sinistro em questão. No que toca às oficinas, todos os sinistrados escolheram as suas oficinas e nenhuma das escolhidas foi associada a fraude anteriores, sendo que também

na totalidade dos casos a oficina não tinha sido seleccionada em casos anteriores. Para finalizar, os sinistrados são do género masculino portugueses na sua maioria, 70% e 79% respectivamente, com pouca informação sobre a sua ocupação, esta apenas é conhecida para 3% dos casos. (Anexos: Tabelas 27 – 51, Gráficos: 21 - 40)

Já o segmento 3, é caracterizado por a apólice ter uma duração média de um ano desde que foi efectuada até à ocorrência do sinistro, pertencendo principalmente ao intervalo dos 25% mais baixo do valor seguro total, 78%, sendo este composto principalmente por veículos ligeiros mas com uma percentagem bastante significativa, cerca de 19%, de veículos de mercadorias, segmentados uniformemente pelas diversas gamas de marcas automóveis, com pouca informação sobre o seu valor de mercado, e uma idade média de cerca de 4 anos. Não existe informação sobre qual o período temporal de quando ocorreram os mesmos, apenas é descrito que ocorreram em Portugal Continental, quase totalmente ao dia de semana, 78.7%, onde 36.2% dos mesmos na região da grande Lisboa e 10.6% na região do grande Porto, sem existir qualquer intervenção das autoridades, onde 19.1% dos mesmos foram registados danos corporais, sem a existência de testemunhas e com uma média de comunicação à seguradora de cerca de 4 dias. Relativamente às oficinas, em todos os casos os sinistrados escolheram a oficina e em qualquer das escolhidas foi associada a fraudes anteriores. Na caracterização dos sinistrados volta-se a verificar uma predominância de indivíduos do género masculino, 72% do total, com nacionalidade portuguesa, 30%, no entanto existo cerca de 64% dos mesmos onde não existe informação sobre qual a sua nacionalidade, um quase total desconhecimento sobre qual a sua ocupação, 97%, assim como de alguma lesão contraída durante o sinistro. (Anexos: Tabelas 52 – 77, Gráficos: 41 - 60)

Por fim, o segmento 4 possui uma média de duração da apólice até ao acidente de cerca de 18 meses, pertencente ao intervalo dos 25% do valor seguro total pela apólice. Os veículos em questão são pertencentes à classe ligeiros, cerca de 77%, e à classe mercadorias com 21%, com predominâncias das marcas Mercedes, Opel, Toyota, Honda e Peugeot e uma idade média de 5 anos. Também não existe informação sobre qual foi a hora aos quais os sinistros ocorreram, mas existem informação que estes ocorreram principalmente em Portugal, nomeadamente na grande Lisboa e no grande Porto fazendo 34.5% do total, principalmente durante os dias de semana, cerca de 71.6% dos casos, sem

qualquer intervenção por parte das autoridades, onde apenas existiu danos corporais em cerca de 12.3% dos sinistros, não existiram quaisquer testemunhas e o reporte à seguradora foi feito numa média de 6 dias. Relativamente às oficinas, todos os segurados escolheram qual a sua oficina e 6% destas tinham sido identificadas como fraudulentas em sinistros anteriores. Sobre o sinistrado, volta-se a verificar a predominância do género masculino de nacionalidade portuguesa, 70.9% e 36.36% respectivamente, contudo volta-se a verificar pouca informação sobre esta variável e um total desconhecimento sobre a sua ocupação. (Anexos: Tabelas 78 – 102, Gráficos: 61 - 80).

### ***b) Criação de matriz de semelhança***

Para a criação da rede foi necessário analisar o comportamento de cada segmento para poder construir as possíveis relações que estes possuem entre eles. Para isso foi criada uma matriz de semelhança que permita relacionar comportamentos semelhantes de forma a quê seja possível uma identificação mais imediata dos valores comuns, será utilizada a média para as variáveis numéricas e a frequência para as variáveis caracter, sendo que para cada variável não única será criada uma matriz de semelhança.

Assim, após a criação das matrizes é possível verificar que a variável que possuem relações entre segmentos é a antiguidade da apólice original em meses (**NUM\_Antig\_Act\_Policy**), cuja média dos valores é aproximadamente comum entre os segmentos 2 e 4, ou seja, em ambos o segmentos, a média para o número de meses desde o a criação apólice até à ocorrência do sinistro é cerca de 18 meses.

<b>NUM_Antig_Act_Policy</b>	<b>Segmento 1</b>	<b>Segmento 2</b>	<b>Segmento 3</b>	<b>Segmento 4</b>
<b>Segmento 1</b>	-	0	0	0
<b>Segmento 2</b>	0	-	0	1
<b>Segmento 3</b>	0	0	-	0
<b>Segmento 4</b>	0	1	0	-

Matriz 1 – Relação da variável **NUM\_Antig\_Act\_Policy** entre segmentos.

A segunda variável que aparece é a antiguidade da última alteração nas coberturas da apólice (**NUM\_Antig\_Alt\_Policy**), sendo que a sua média é semelhante entre os segmentos 3 e 4 principalmente, com valores à volta dos 163 dias.

<b>NUM_Antig_Alt_Policy</b>	<b>Segmento 1</b>	<b>Segmento 2</b>	<b>Segmento 3</b>	<b>Segmento 4</b>
<b>Segmento 1</b>	-	0	0	0
<b>Segmento 2</b>	0	-	0	0
<b>Segmento 3</b>	0	0	-	1
<b>Segmento 4</b>	0	0	1	-

Matriz 2 – Relação da variável *NUM\_Antig\_Alt\_Policy* entre segmentos.

De seguida surge a variável que revela o valor seguro do veículo (**VAL\_Prop\_Exposure**), demonstrando uma relação dos segmentos 2 e 3, com valores semelhantes a rondar os 3400€.

<b>VAL_Prop_Exposure</b>	<b>Segmento 1</b>	<b>Segmento 2</b>	<b>Segmento 3</b>	<b>Segmento 4</b>
<b>Segmento 1</b>	-	0	0	0
<b>Segmento 2</b>	0	-	1	0
<b>Segmento 3</b>	0	1	-	0
<b>Segmento 4</b>	0	0	0	-

Matriz 3 – Relação da variável *VAL\_Prop\_Exposure* entre segmentos.

Verifica-se também valores semelhantes entre segmentos na variável que revela a hora do sinistro (**COD\_Claim\_Time**), nomeadamente entre os segmentos 1, 3 e 4 nomeadamente no valor S – Sem informação, da hora de ocorrência do sinistro.

<b>COD_Claim_Time</b>	<b>Segmento 1</b>	<b>Segmento 2</b>	<b>Segmento 3</b>	<b>Segmento 4</b>
<b>Segmento 1</b>	-	0	1	1
<b>Segmento 2</b>	0	-	0	0
<b>Segmento 3</b>	1	0	-	1
<b>Segmento 4</b>	1	0	1	-

Matriz 4 – Relação da variável *COD\_Claim\_Time* entre segmentos.

A variável que se segue é o número de anos de antiguidade do veículo 1 ao ano de Construção (**NUM\_V1\_Age\_FR**), com valores semelhantes entre os segmentos 3 e 4 referindo uma antiguidade do veículo 1 do sinistro entre os 4 e 5 anos.

<b>NUM_V1_Age_FR</b>	<b>Segmento 1</b>	<b>Segmento 2</b>	<b>Segmento 3</b>	<b>Segmento 4</b>
<b>Segmento 1</b>	-	0	0	0
<b>Segmento 2</b>	0	-	0	0
<b>Segmento 3</b>	0	0	-	1
<b>Segmento 4</b>	0	0	1	-

Matriz 5 – Relação da variável *NUM\_V1\_Age\_FR* entre segmentos.

Por fim, também a variável que dita a Nacionalidade do segurado (**COD\_NAT\_CI**) é representada com relações entre segmentos, mais propriamente a ligação entre os segmentos 1 e 2 e a ligação entre os 3 e 4 mais propriamente ao nível da frequência de

segurados portugueses nos segmentos 1 e 2, e o desconhecimento da nacionalidade da grande maior parte dos segurados nos segmentos 3 e 4.

<b>COD_NAT_CI</b>	<b>Segmento 1</b>	<b>Segmento 2</b>	<b>Segmento 3</b>	<b>Segmento 4</b>
<b>Segmento 1</b>	-	1	0	0
<b>Segmento 2</b>	1	-	0	0
<b>Segmento 3</b>	0	0	-	1
<b>Segmento 4</b>	0	0	1	-

Matriz 6 – Relação da variável *COD\_NAT\_CI* entre segmentos.

Assim, fazendo uma média simples das matrizes apresentadas em cima, verifica-se que os segmentos 3 e 4 são os que aparecem relacionados mais vezes, em 66,67% dos casos.

<b>% Relações</b>	<b>Segmento 1</b>	<b>Segmento 2</b>	<b>Segmento 3</b>	<b>Segmento 4</b>
<b>Segmento 1</b>	-	16.67%	0.00%	0.00%
<b>Segmento 2</b>	16.67%	-	16.67%	16.67%
<b>Segmento 3</b>	0.00%	16.67%	-	66.67%
<b>Segmento 4</b>	0.00%	16.67%	66.67%	-

Matriz 7 – Percentagem das relações mais frequentes entre segmentos.

As restantes variáveis não referidas anteriormente possuem duas características, ou obtêm valores distintos em todos os segmentos, como a variável referente à natureza da lesão no segurado (*COD\_INJ\_Nature\_C*), as variáveis relacionadas com a potência dos veículos (*ENGINE*) e as variáveis relacionadas com os quilómetros dos veículos no sinistro (*KM*), ou possuem valores os mesmos valores mais frequentes ou com média semelhante para todos os segmentos, exemplo as variáveis que referem se o Sinistro ocorreu ou não num Final de Semana ou Feriado (*FLG\_Claim\_Wkend*), qual a NUT 3 onde se deu o acidente (*COD\_Claim\_Loc\_Nuts*), se o sinistro ocorreu no Estrangeiro ou Portugal (*FLG\_Claim\_Overseas*), se o tomador é reclamante (*FLG\_Claim\_Insured*), se existiu ou não intervenção das autoridades (*FLG\_Claim\_Police\_Rep*), se existiu ou não danos corporais (*FLG\_Claim\_Injured*), qual o género do segurado (*COD\_Gender\_CI*), se o cliente escolheu a oficina (*FLG\_VI\_Chose\_Garage*), se a oficina associada ao veículo segurado teve fraudes anteriores (*FLG\_VI\_Repair\_PFraud*) e se os sinistros anteriores envolvendo a oficina do veículo 1 (*FLG\_PC\_V1\_Provider*). Os resultados destas podem ser encontrados em anexo.

## V. Conclusões

Após a análise dos segmentos, é possível criar a rede regular, do tipo *small world*, pois esta assenta sobre a dinâmica onde se pretende caracterizar quais os tipos de comportamentos mais prováveis a serem observados na estrutura, que interligue os mesmos pelas variáveis com valores mais frequentes comuns. Posto isto, a seguinte rede de relações é obtida:

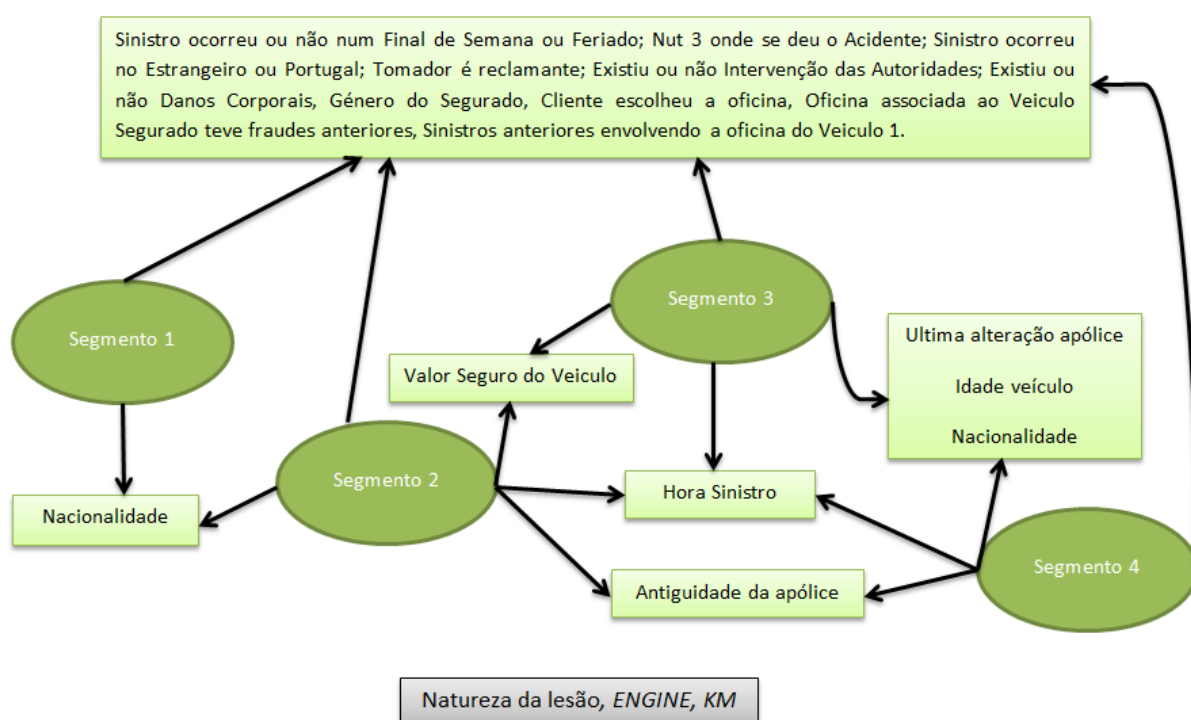


Figura 4 – Materialização da Rede entre segmentos.

Em conclusão, verifica-se que, para a amostra em estudo, as características de se o Sinistro ocorreu ou não num Final de Semana ou Feriado, qual a Nomenclatura de Unidade Territorial (NUT) de nível 3 onde se deu o acidente, se o sinistro ocorreu no estrangeiro ou em Portugal, se o tomador da apólice é reclamante, se existiu ou não Intervenção das Autoridades, se existiram ou não Danos Corporais, o género do Segurado, se o cliente escolheu a oficina, se a oficina associada ao veículo segurado teve fraudes anteriores e os sinistros anteriores envolvendo a oficina do Veiculo 1 não são explicativas de factores de diferenciação na rede pois estas são comuns a todos os segmentos, também nesta situação mas em contraponto, encontram-se as características da natureza da lesão, do motor (*Engine*) e dos quilómetros dos veículos (KM), pois estes obtêm valores diferentes para todos os segmentos.

No que toca a dados concretos, os segmentos 1 e 2 relacionam-se na medida em que na sua maioria, a nacionalidade do tomador de seguros é portuguesa, no entanto, esta relaciona também os segmentos 3 e 4 mas por esta se encontrar preenchida como “desconhecido”.

O valor seguro pela apólice relaciona os segmentos 2 e 3, ou seja, os casos para os quais os clientes apenas foram seleccionados ou como fraudulentos ou apenas como suspeitos, apresentam valores semelhantes, cerca de 11000€. Em ambos os casos onde se registam casos de fraude confirmada, segmento 2 e 4, a média da antiguidade da apólice é a mesma, 18 meses, também juntando o segmento 3 aos dois anteriores não existe informação sobre a hora de ocorrência do sinistro tanto para os casos confirmados de fraude como para os casos apenas suspeitos.

A ligar os casos apenas suspeitos, segmento 3, com os casos suspeitos e confirmados, segmento 4, está a ultima alteração às coberturas da apólice foi efectuada com o mesmo número médio de dias, cerca de 111, assim como a idade do automóvel coberto, 4/5 anos, pela apólice e a nacionalidade “desconhecida” dos clientes. Neste ultimo caso, não é perceptível qual diferença entre significativa entre o segmento 3 e 4 questionando assim como são aplicadas as técnicas de classificação por parte dos criadores da amostra, pois para além dos valores comuns entre as variáveis que interceptam estes, as variáveis que demonstram independência de qualquer dos segmentos, obtém percentagem de valores omissos bastantes elevados, na ordem dos 90%.

Na perspectiva deste investigador, é possível a aplicação deste tipo de metodologia a outros casos da indústria seguradora no caso dos seguros automóveis, devido à facilidade de implementação deste tipo de redes e na rapidez na elaboração e interpretação dos resultados assim obtidos. Contudo, é necessário que a amostra em questão contenha pelo menos, as variáveis que foram seleccionadas para o estudo desta amostra, e principalmente, as variáveis indicadores dos casos de suspeita de fraude e os casos de fraude confirmada. Também neste aspecto, é necessário que a amostra onde será implementada a rede, contenha um número de registo significativo para resultados mais significativos.

## **VI. Limitações e investigação futura**

A primeira limitação prende-se com o facto da obtenção de dados. Devido às burocracias e questões éticas por parte dos detentores da amostra sendo que, a facultação da mesma foi um processo moroso o que levou a que o arranque do projecto se iniciasse alguns meses depois do planeado.

Outra limitação é o número de observações da amostra, um número de apenas 355.483 observações é limitado para uma análise modelatória desta dimensão.

Um dos pontos que deverá ser abordado também é o facto de a amostra estar mal preenchida, ou seja neste caso conter um número elevado de valores a vazio, não só nos valores para as variáveis escolhidas como para a totalidade da amostra. Focando as variáveis escolhidas para análise, exemplificando com a tabela 102 do anexo para a totalidade da amostra. Como é possível verificar, existem bastantes variáveis seleccionadas para o estudo da amostra cujas quais tem uma percentagem de valores a vazio a rondar os 90%. E este problema densifica-se à medida que se alonga pela análise dos segmentos, é possível ver nas tabelas 103, 104, 105 e 106 do anexo, existem variáveis rondar os 100%, nomeadamente nos segmentos 2, 3 e 4 enviesando assim a análise efectuada. Exemplo disso são as variáveis relacionadas com a potência dos veículos, assim como os quilómetros dos mesmos envolvidos no sinistro.

Por fim nas limitações, não existe conhecimento de como forma foram indicado como casos suspeitos e/ou fraudulentos por parte dos criadores da amostra, limitando assim as conclusões que se possam retirar desta.

No que toca a estudos futuros, uma análise de correlações assim como uma análise factorial entre as variáveis para assim, estudar as variações de valores entre as várias variáveis constituintes da amostra, como também a análise de possíveis relações espúrias<sup>35</sup> entre estas.

---

<sup>35</sup> Uma relação espúria significa uma falsa relação, seja esta positiva ou negativa entre duas variáveis, esta pode ter origem na pura coincidência entre as duas variáveis ou que a mesma seja criada por uma terceira variável.



A utilização de técnicas de segmentação de amostras como a criação de *clusters* utilizando técnicas dos k-vizinhos mais próximos ou os *self-organizing maps*, também conhecidos como SOM, permitem uma análise *à priori* da amostra, desconsiderando já as variáveis de suspeição e confirmação de casos de fraude que podem ser posteriormente cruzadas para analisar os casos de *miss-classification* por parte das regras de negócio e da análise de *outliers*, efectuadas pela criadores da apólice.

## VII. Bibliografia

ACL (2013). *Fraud Detection Using Data Analytics in the Insurance Industry*

Albert R. & Barabási A. (2002). *Statistical mechanics of complex networks*.

Araújo, T. V. (2006). Redes em Economia: criação de estruturas e auto-organização em sistemas económicos complexos.

Artís, M., Ayuso, M., Guillé, M. (2002). *Detection of Automobile Insurance Fraude With Discrete Choise Models and Misclassified Claims*.

Baker, T. (1996). *On the Genealogy of Moral Hazard*.

Baldock, T. (1997). *Insurance Fraud*.

Blass, T (2002). *The Man Who Shocked the World: The Life and Legacy of Stanley Milgram*.

Brites, J. C. (2006). Fraude em Seguros. Bolsa dos Seguros: Revista de Seguros de Pensões, 22 (9), pp. 20-27.

Cialdini, R. B. (1993). *Influence: The Psychology of Persuasion*.

Clarke, M. (1989). *Insurance Fraud*.

Clarke, M. (1990). *The control of insurance fraud: a comparative view*.

Correia, A. (2007). Fraude, um assunto que a todos diz respeito! APS Notícias.

Correia, A. (2010). Fraude: Introdução. In F. Gilberto, Os Grandes Desafios da Indústria Seguradora.

Costa, A. M. A. (1999). Capítulo III, Dos Crimes Contra o Património em Geral. In J. Figueiredo Dias (dir.), Comentário Conimbricense do Código Penal.

Cummins, D. & Tennyson, S. (1996). *Moral Hazard in insurance claiming: Evidence from automobile insurance*.

Dean, D. H. (2004). *Perceptions of the Ethicality of Consumer Insurance Claim Fraud*.

Decreto-Lei nº 408/1979 de 25 de Setembro

Decreto-Lei 94-B/98, de 17 de Abril

Decreto-Lei nº 291/2007, de 21 de Agosto

Decreto-Lei nº72/2008 de 16 de Abril.

Decreto-Lei 2/2009, de 5 de Janeiro

Derrig, Richard A. (2002). *Insurance Fraud*.

Derrig, Richard A.; Johnston, Daniel J.; Sprinkel, Elizabeth A.(2006). *Auto Insurance Fraud: Measurements and Effords to combat it*.

Diario da Republica nº170, Série I.

Dionne, G. & Gagné, R. (2002). *Replacement Cost Endorsement and Opportunistic Fraud in Automobile Insurance*.

Dionne, G.; Wang, K. S. (2013). *Does insurance fraud in automobile theft insurance fluctuate with the business cycle?*

Erin Insurance (2012). *Erie Insurance Improves Its Detection Process and Stays True to Its Values with LexisNexis® FraudFocus Enhanced*

Fachin, Odília (2001). Fundamentos de metodologia. São Paulo: saraiva.

Fidel, Raya (1992). *The case study method: a case study*. GLAZIER, Jack D. & POWELL, Ronald R. *Qualitative research in information management*. Englewood, CO: Libraries Unlimited. Pp.37-50.

GNR (2013). Fraude em seguros: Colaboração entre a GNR e a APS.

Instituto de Seguros de Portugal (2011). Seguro Automóvel. In Instituto de Seguros de Portugal, Guia de Seguros e Fundos de Pensões.

Insurance Research Council (1991). *Public Attitude Monitor*.

Insurance Research Council (1993). *Public Attitude Monitor*.

Lang, M; Wambach, A. (2013). *The fog of fraud – Mitigating fraud by strategic ambiguity*.

Lesch, William C.; Brinkmann, Johannes (2011). *Consumer Insurance Fraud/Abuse as Co-creation and Co-responsibility: A New Paradigm*.

- Lincoln, R., Wells, H. & Petherick, W. (2003). *An Exploration of Automobile Insurance Fraud*.
- Martins, João Valente (2006). Notas práticas sobre o contrato de seguro.
- Mendes, R. V. (1998a). Medidas de complexidade e auto-organização.
- Mendes, R. V. (1998b). *Conditional exponents, entropies and a measure of dynamical self-organization*.
- Mendes, R. V. (2006). Complexidade e auto-organização no mundo natural: dos procariontes à globalização.
- Miyazaki, A. D. (2008). *Perceived Ethicality of Insurance Claim Fraud: Do Higher Deductibles Lead to Lower Ethical Standards?*
- Newman, M. E. J. (2003). *The structure and function of complex networks*.
- Niemi, Hannu (1995). *Insurance Fraud*.
- SAS (2012). *Combating Insurance Claims Fraud. How to Recognize and Reduce Opportunistic and Organized Claims Fraud*.
- Sparrow, M., (1996). *License to Steal*.
- Tennyson, S. (1997). *Economic institutions and individual ethics: A study of consumer attitudes toward insurance fraud*.
- Tennyson, S. (2002). *Insurance Experience and Consumers' Attitudes Toward Insurance Fraud*.
- Tennyson, S. (2008). *Moral, Social, and Economic Dimensions of Insurance Claims Fraud*.
- Viaene, S.; Dedene, G. (2004). *Insurance Fraud: Issues and Challenges*.
- Watts, D. (1999) *Small Worlds*.
- Yin, R. (1994). *Case Study Research: Design and Methods*. Thousand Oaks, CA: SAGE Publications

## VIII. Anexos

### a) Gráficos

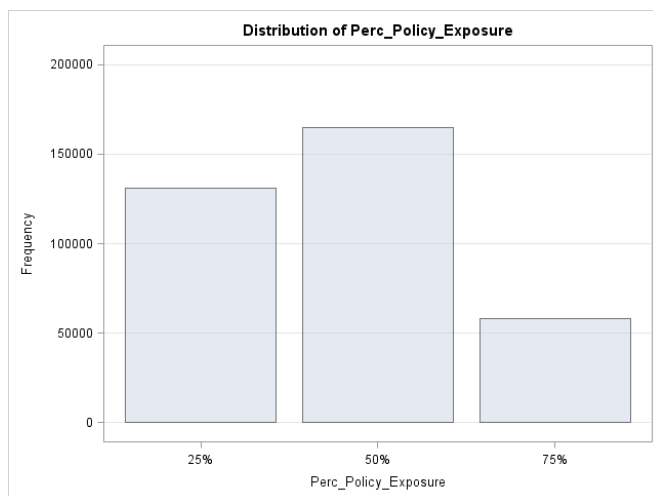


Gráfico 1 – Localização das apólices consoante o quartil de percentagem do valor seguro pela apólice para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

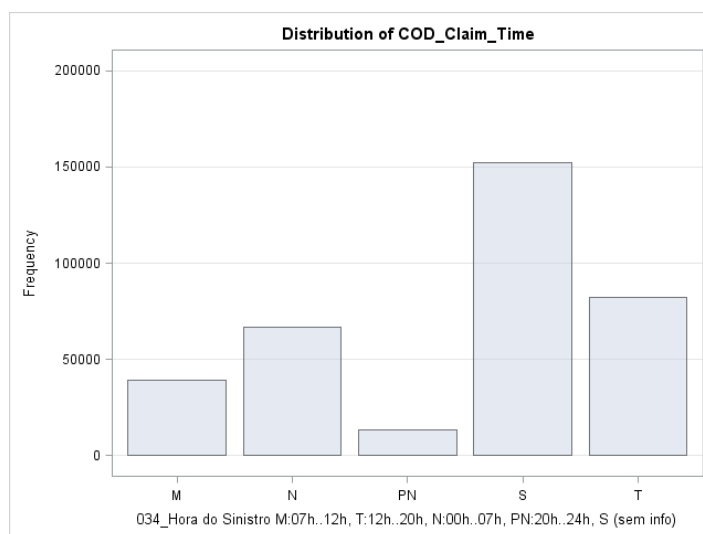


Gráfico 2 – Distribuição temporal da ocorrência dos sinistros para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

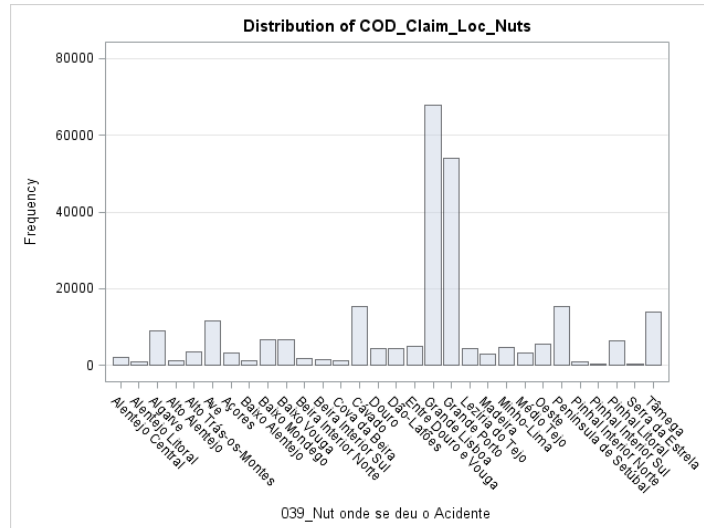


Gráfico 3 – Distribuição geográfica por Nut3 da ocorrência dos sinistros para os casos onde *FLG\_Susp\_Fraud* = 0 e *FLG\_Conf\_Fraud* = 0.

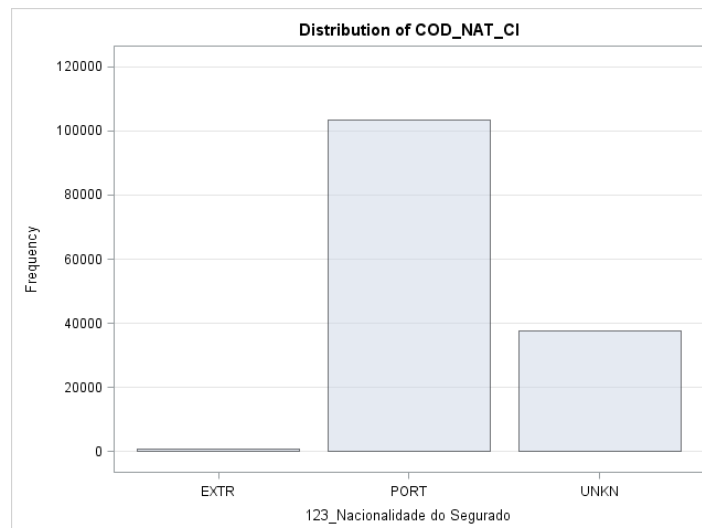


Gráfico 4 – Nacionalidade do sinistrado para os casos onde *FLG\_Susp\_Fraud* = 0 e *FLG\_Conf\_Fraud* = 0.

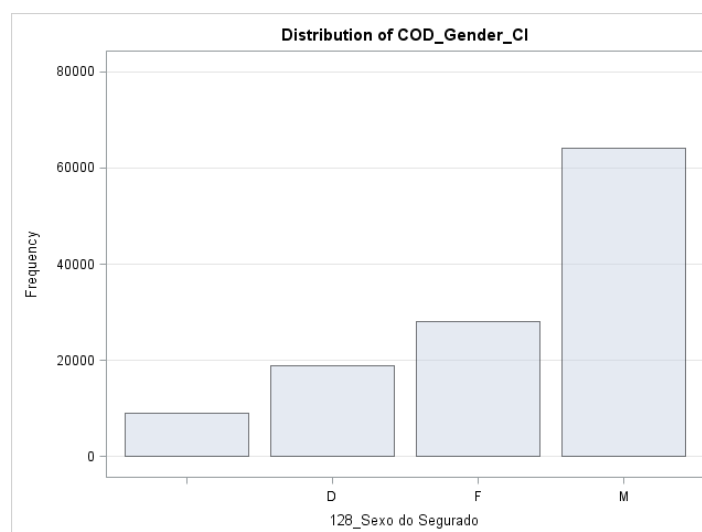


Gráfico 5 – Género do sinistrado para os casos onde *FLG\_Susp\_Fraud* = 0 e *FLG\_Conf\_Fraud* = 0.

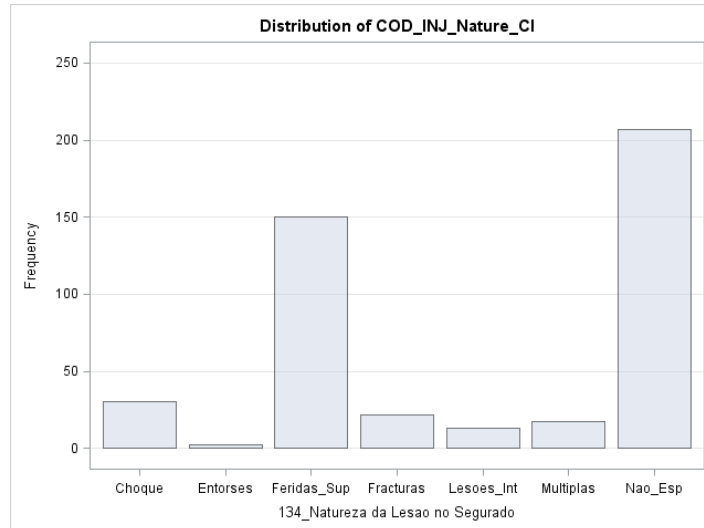


Gráfico 6 – Tipos de lesão decorrente do sinistro para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

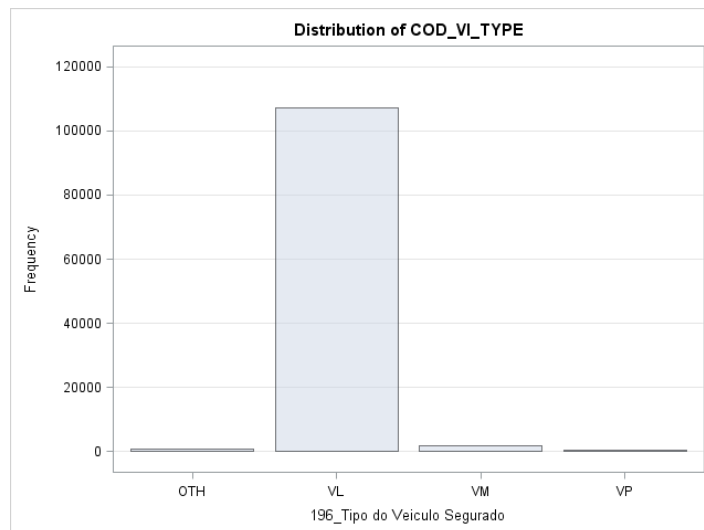


Gráfico 7 – Tipos de veículo seguro para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

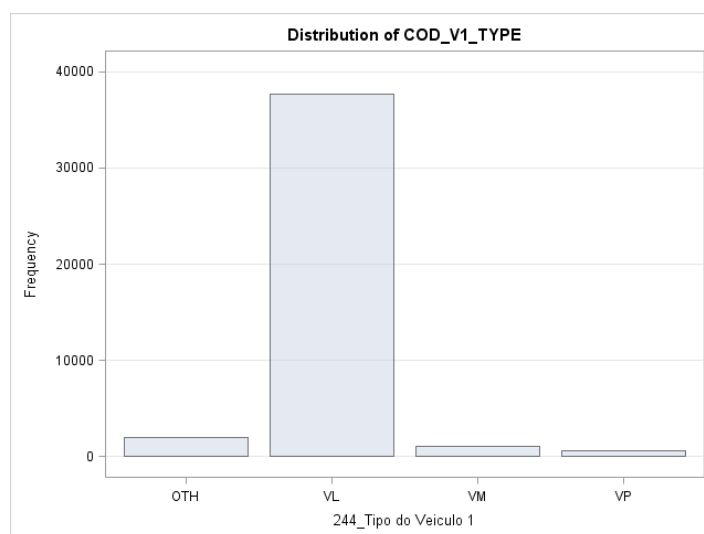


Gráfico 8 – Tipo do veículo 1 no sinistro para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

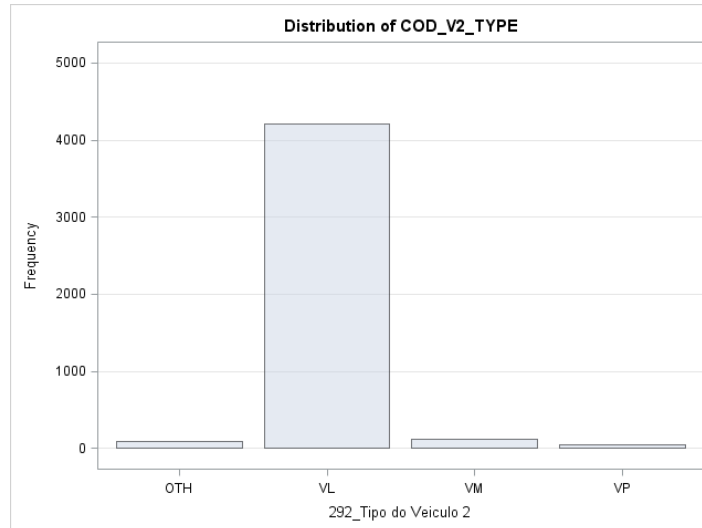


Gráfico 9 – Tipo do veículo 2 no sinistro para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

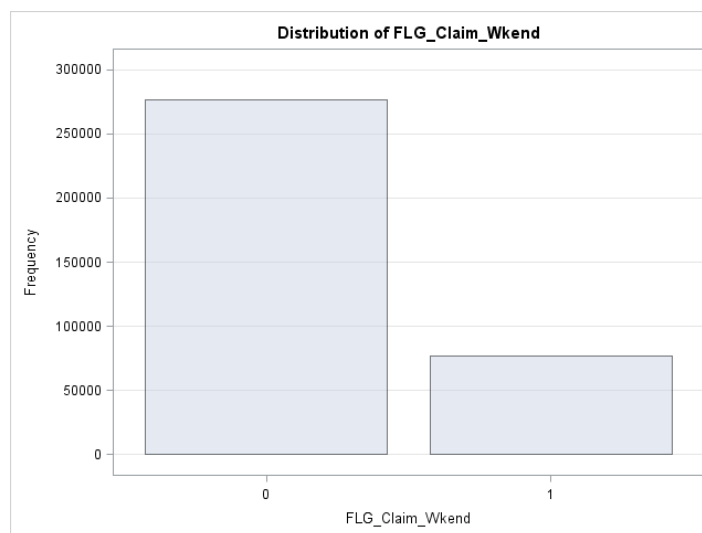


Gráfico 10 – Variável binária indicativa se o acidente ocorreu num dia de semana (0) ou se ocorreu num fim-de-semana/feriado (1) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

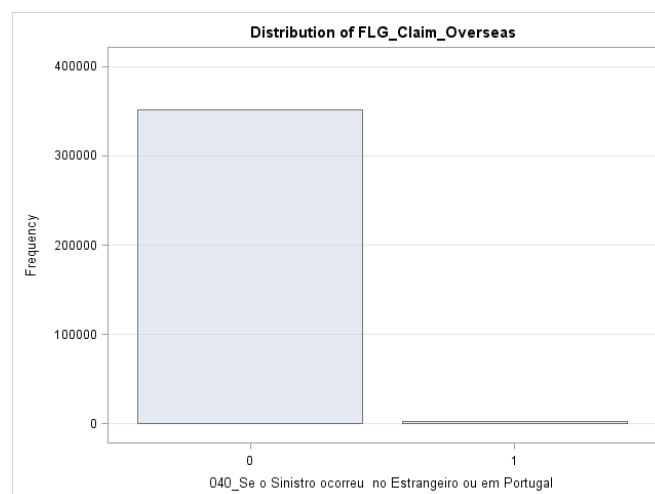


Gráfico 11 – Variável binária indicativa se o acidente ocorreu num em Portugal (0) ou se ocorreu no estrangeiro (1) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .



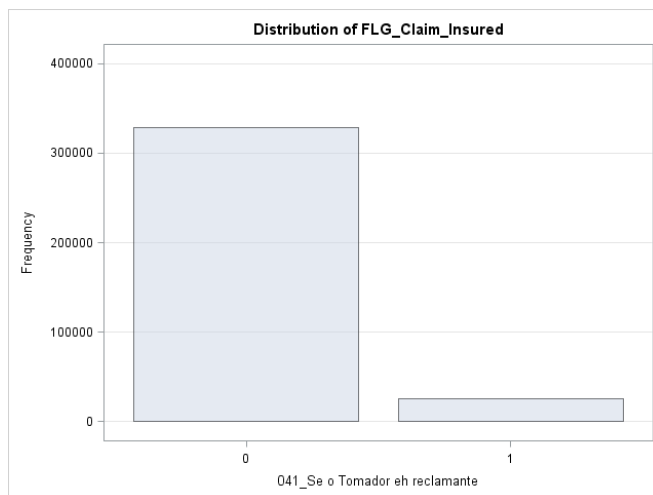


Gráfico 12 – Variável binária indicativa se o tomador da apólice é reclamante (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

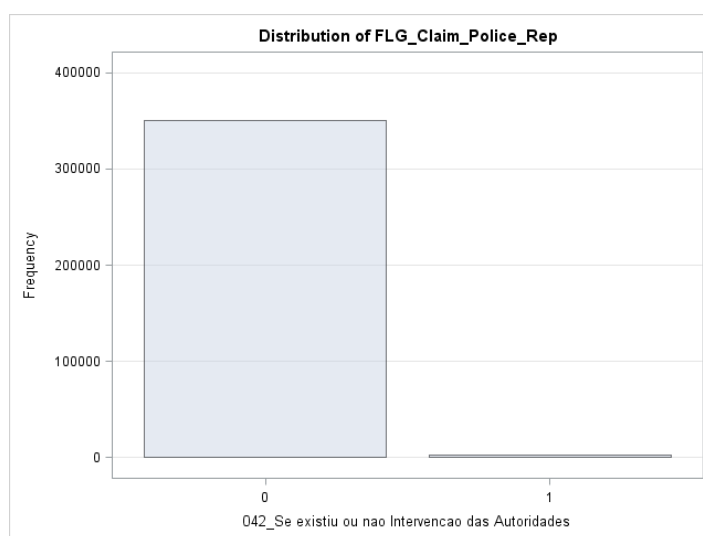


Gráfico 13 – Variável binária indicativa se existiu intervenção das autoridades (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

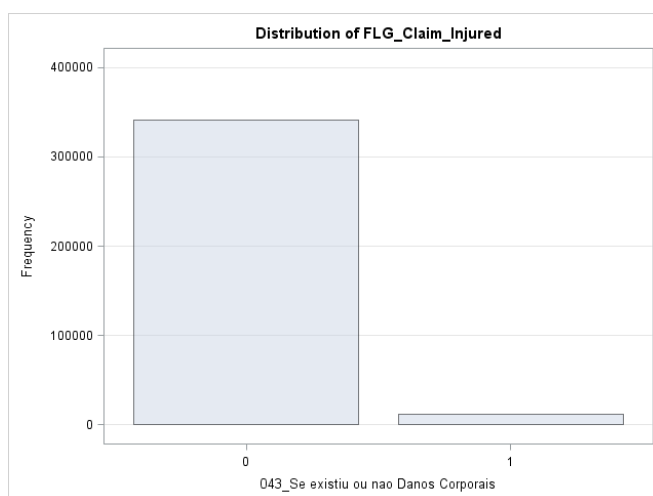


Gráfico 14 – Variável binária indicativa se existiu danos corporais (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

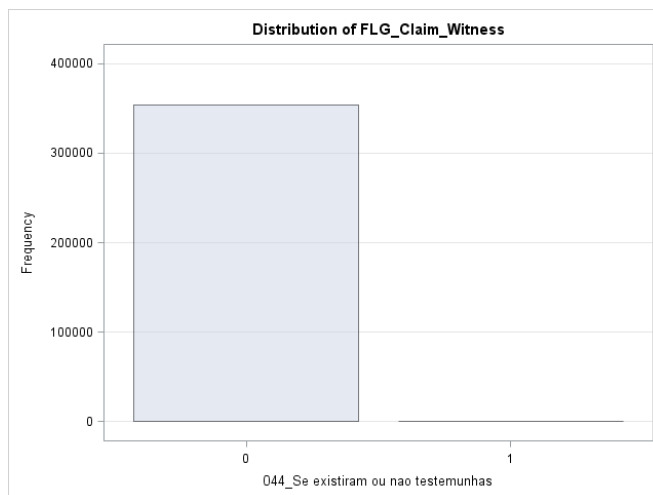


Gráfico 15 – Variável binária indicativa se existiu testemunhas do sinistro (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

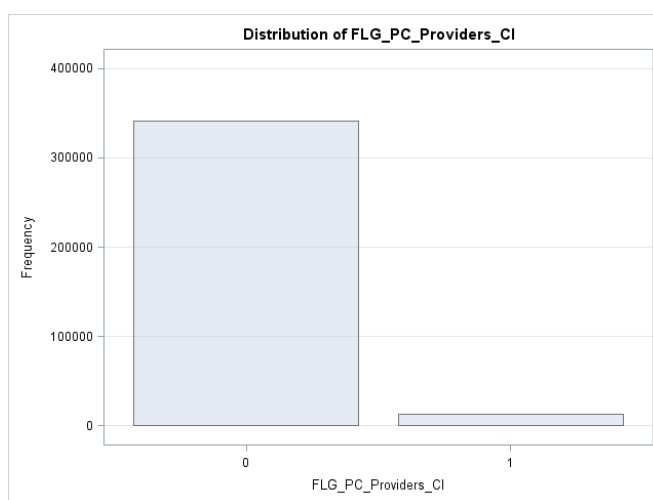


Gráfico 16 – Variável binária que indica se há Sinistros anteriores envolvendo o Segurado e Oficina (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

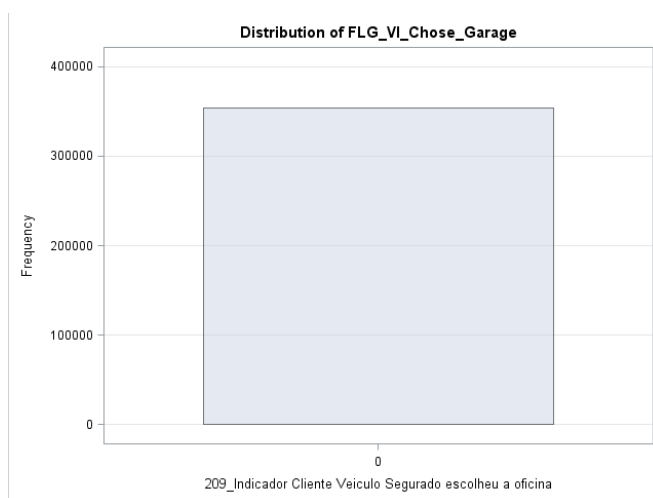


Gráfico 17 – Variável binária que indica se o Segurado escolheu a Oficina (0) ou não (1) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

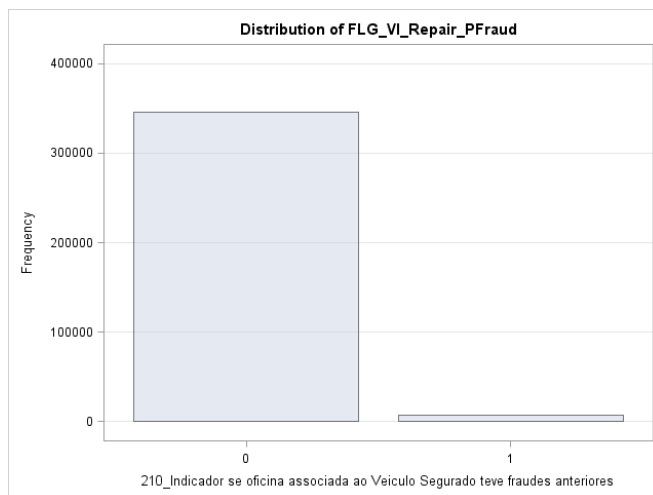


Gráfico 18 – Variável binária que indica se a Oficina está associada a fraude (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

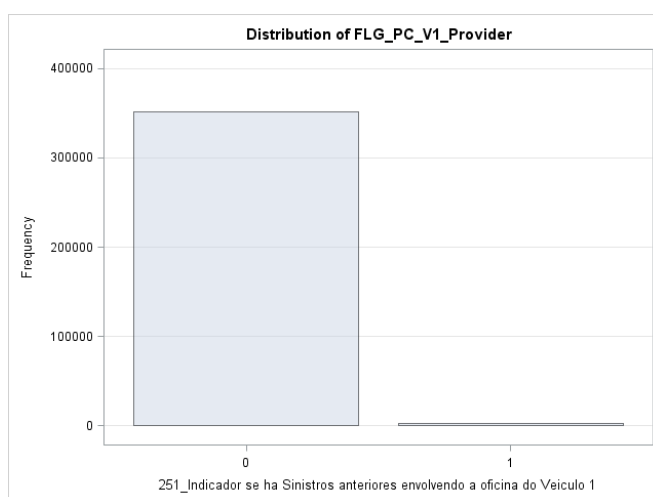


Gráfico 19 – Variável binária se existem sinistros anteriores envolvendo (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

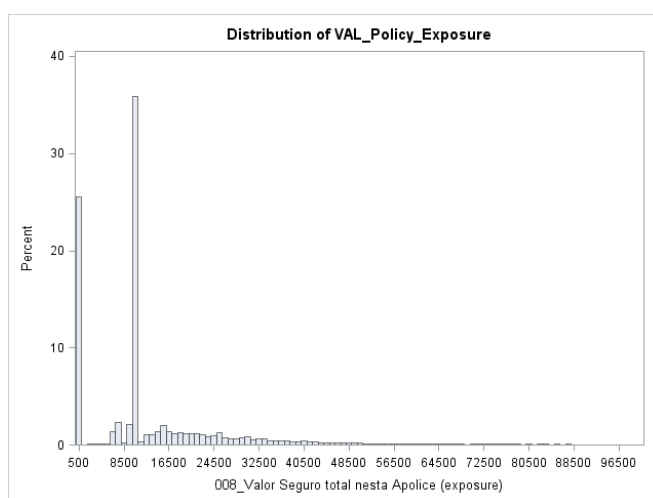


Gráfico 20 – Distribuição do valor seguro pela apólice para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .



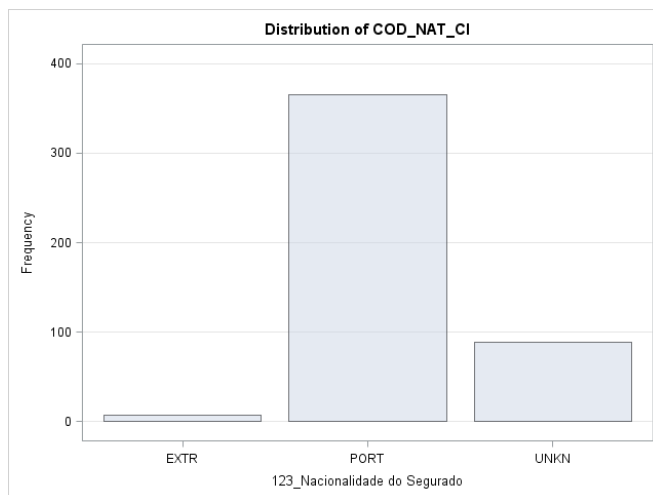


Gráfico 24 – Nacionalidade do sinistrado para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .

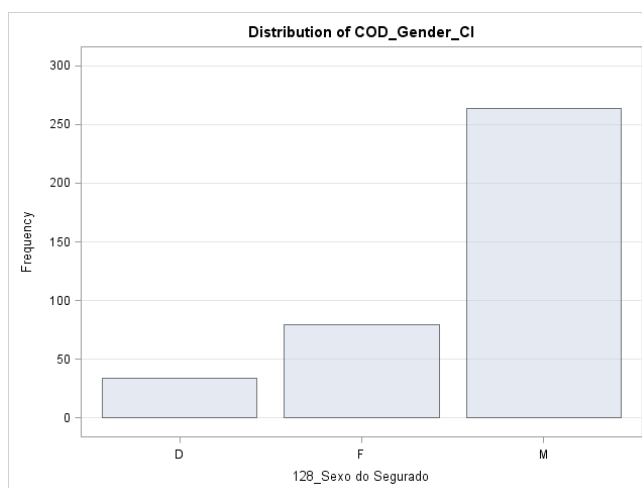


Gráfico 25 – Género do sinistrado para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .

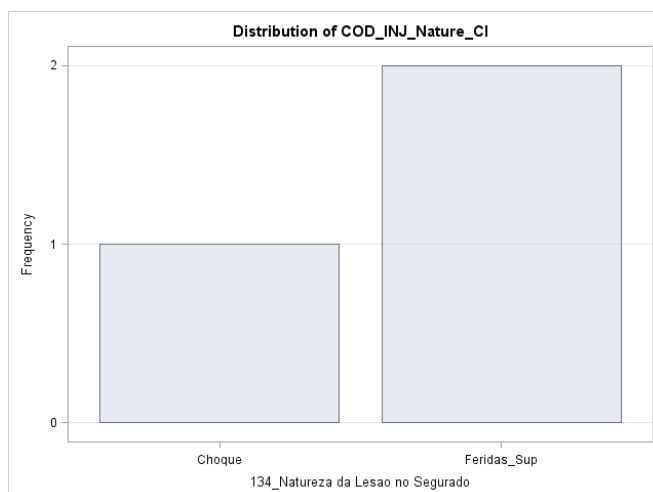


Gráfico 26 – Tipos de lesão decorrente do sinistro para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .

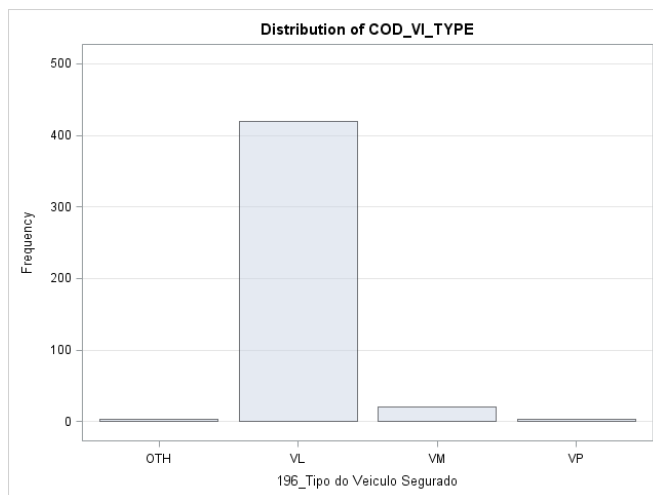


Gráfico 27 – Tipos de veículo seguro para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .

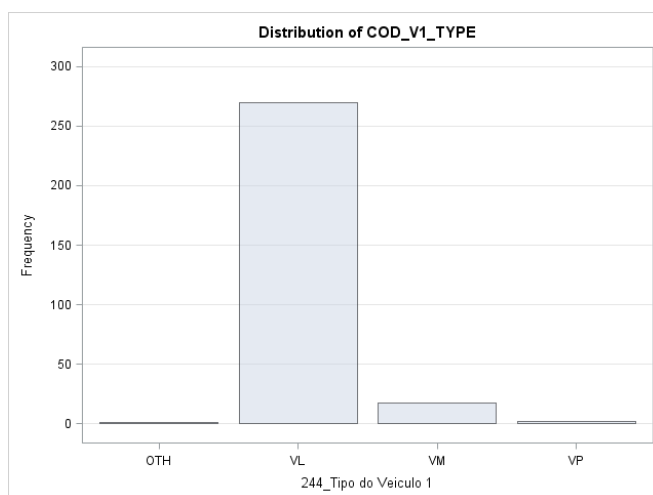


Gráfico 28 – Tipo do veículo 1 no sinistro para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .

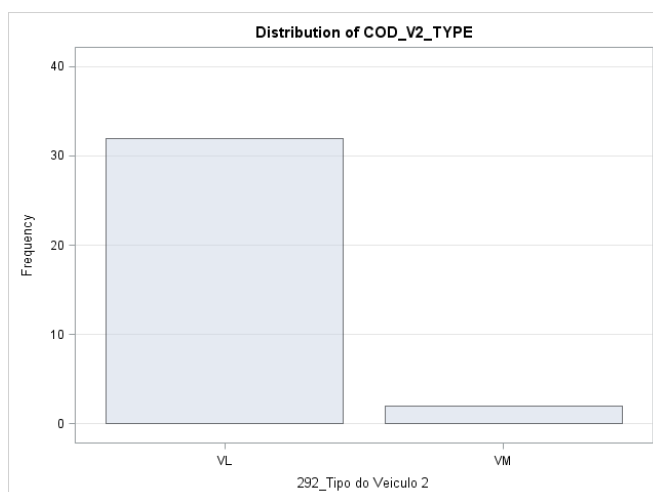


Gráfico 29 – Tipo do veículo 2 no sinistro para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .

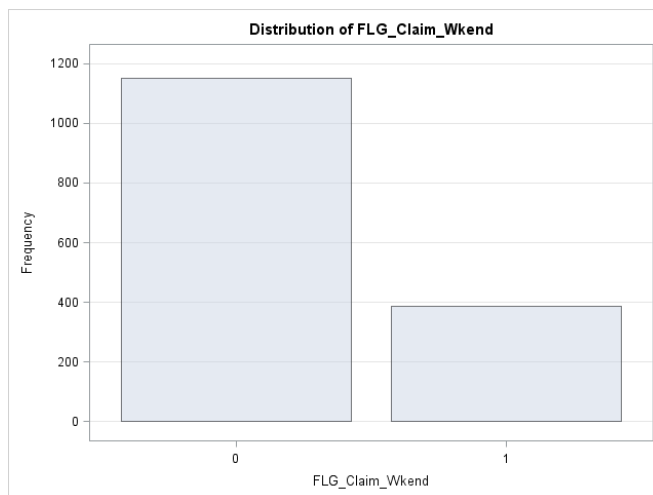


Gráfico 30 – Variável binária indicativa se o acidente ocorreu num dia de semana (0) ou se ocorreu num fim-de-semana/feriado (1) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .

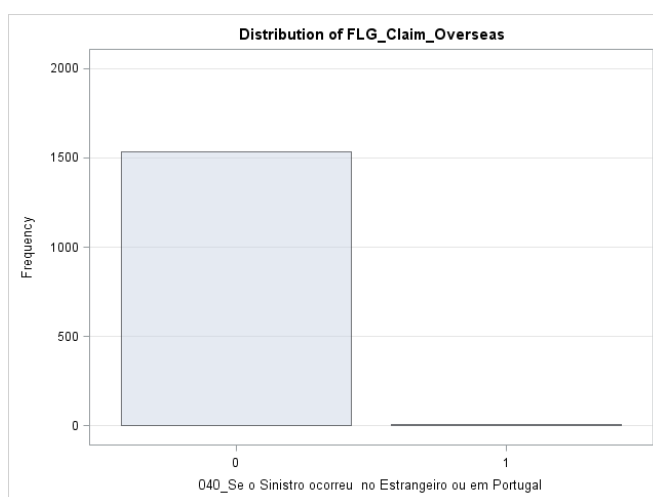


Gráfico 31 – Variável binária indicativa se o acidente ocorreu num em Portugal (0) ou se ocorreu no estrangeiro (1) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .

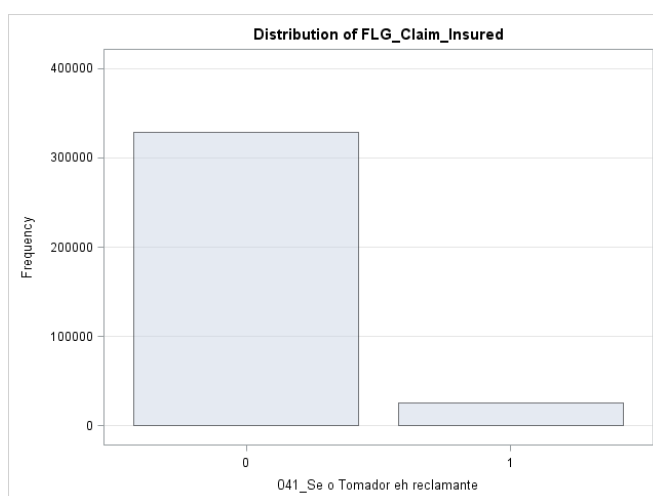


Gráfico 32 – Variável binária indicativa se o tomador da apólice é reclamante (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .

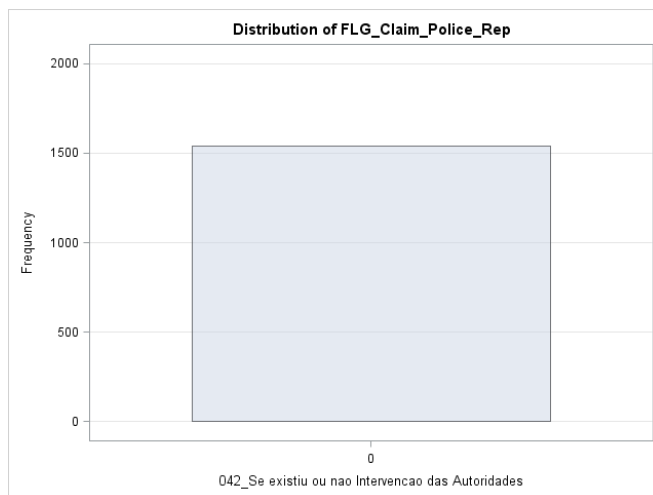


Gráfico 33 – Variável binária indicativa se existiu intervenção das autoridades (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .

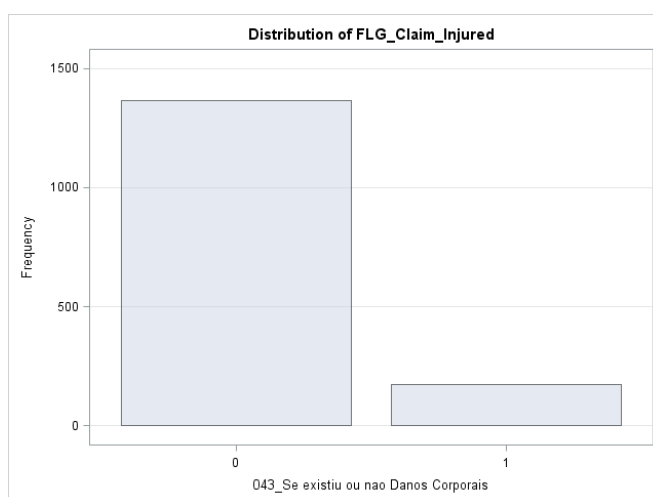


Gráfico 34 – Variável binária indicativa se existiu danos corporais (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .

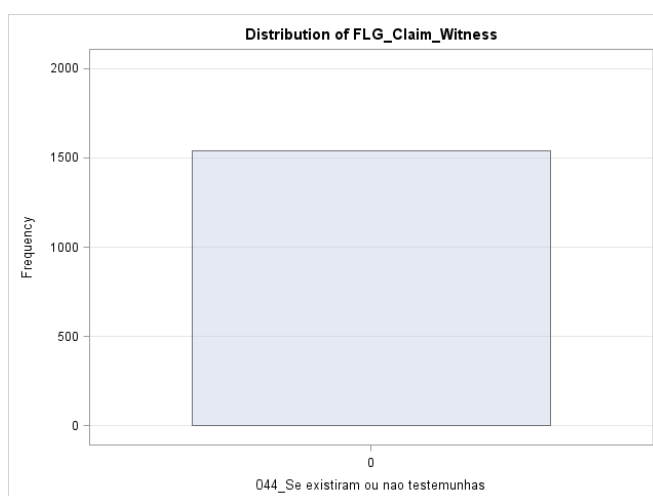


Gráfico 35 – Variável binária indicativa se existiu testemunhas do sinistro (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .



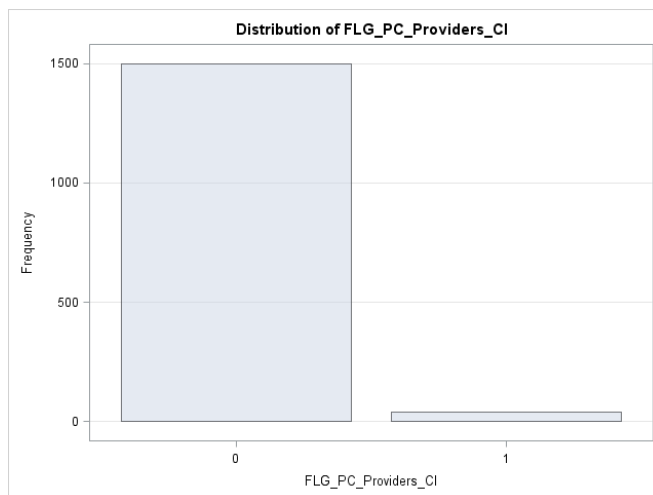


Gráfico 36 – Variável binária que indica se há Sinistros anteriores envolvendo o Segurado e Oficina (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .

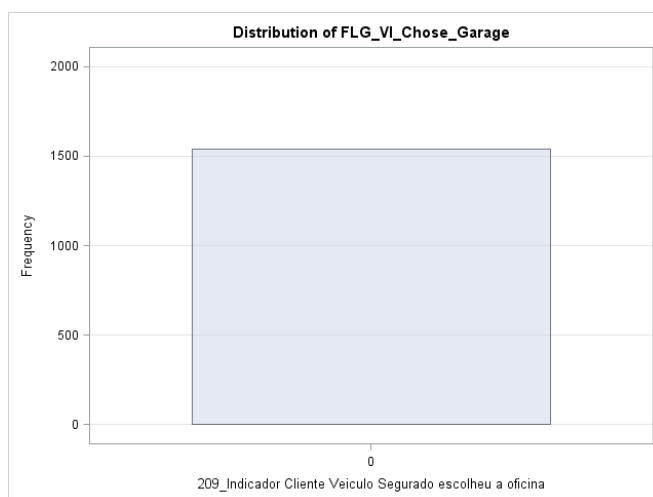


Gráfico 37 – Variável binária que indica se o Segurado escolheu a Oficina (0) ou não (1) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .

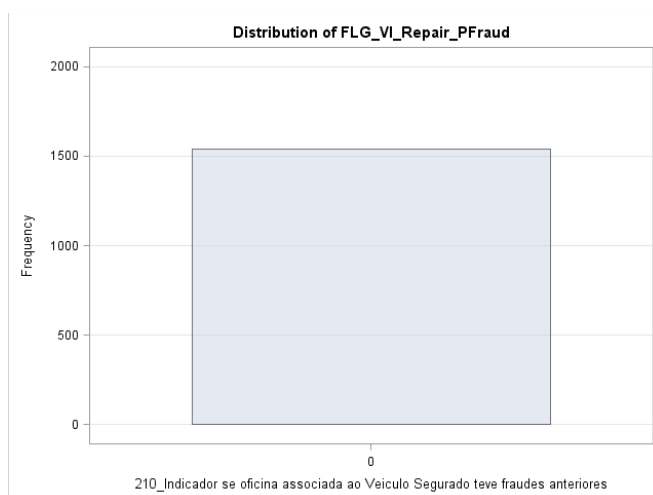


Gráfico 38 – Variável binária que indica se a Oficina está associada a fraude (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .

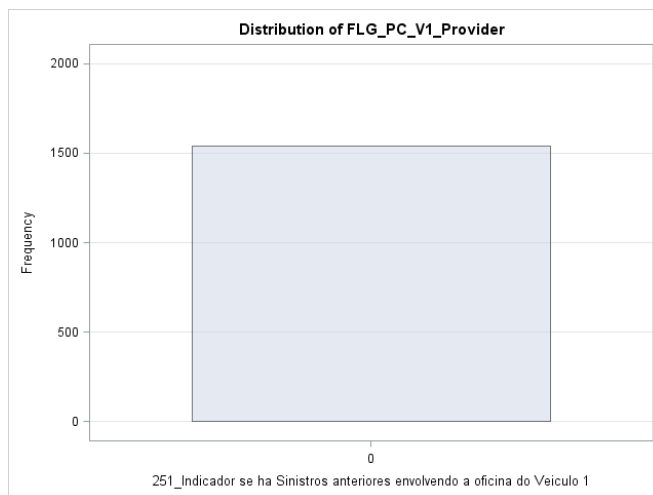


Gráfico 39 – Variável binária se existem sinistros anteriores envolvendo (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .

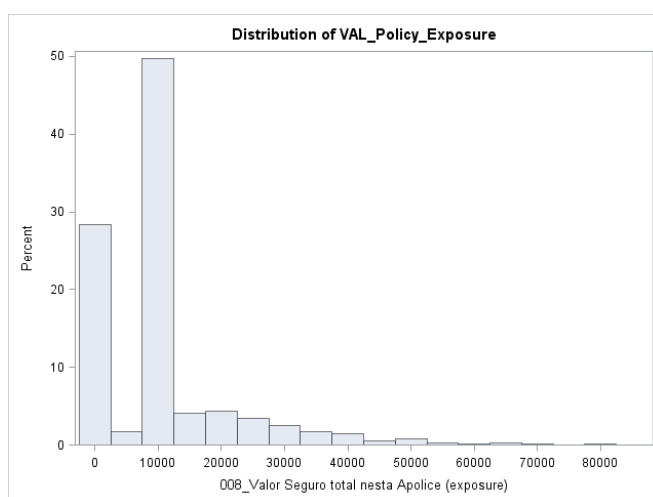


Gráfico 40 – Distribuição do valor seguro pela apólice para os casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .

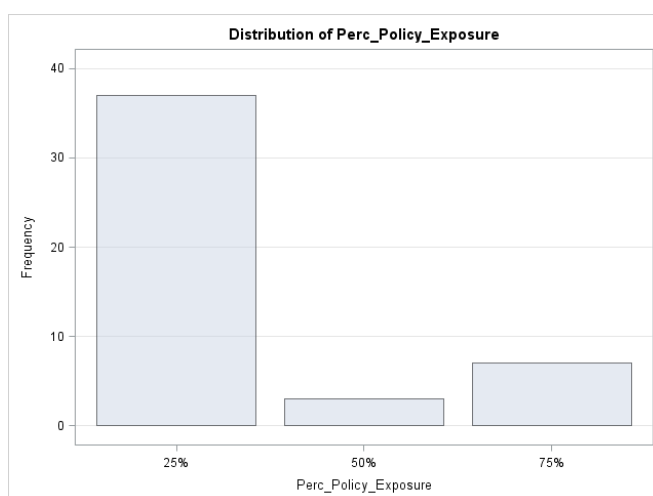


Gráfico 41 – Localização das apólices consoante o quartil de percentagem do valor seguro pela apólice para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

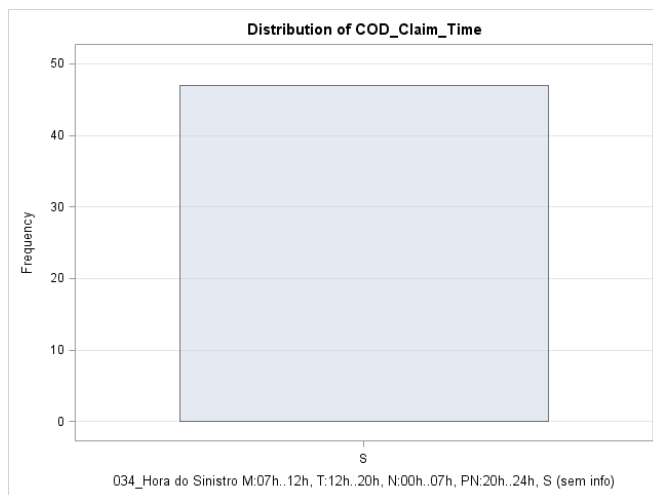


Gráfico 42 – Distribuição temporal da ocorrência dos sinistros para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

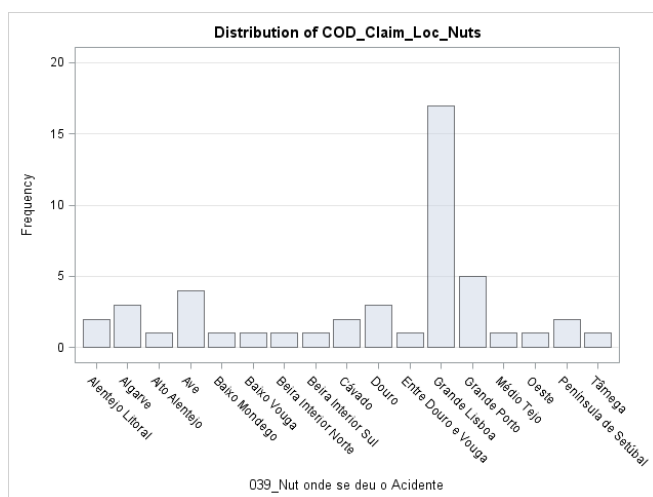


Gráfico 43 – Distribuição geográfica por Nut3 da ocorrência dos sinistros para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

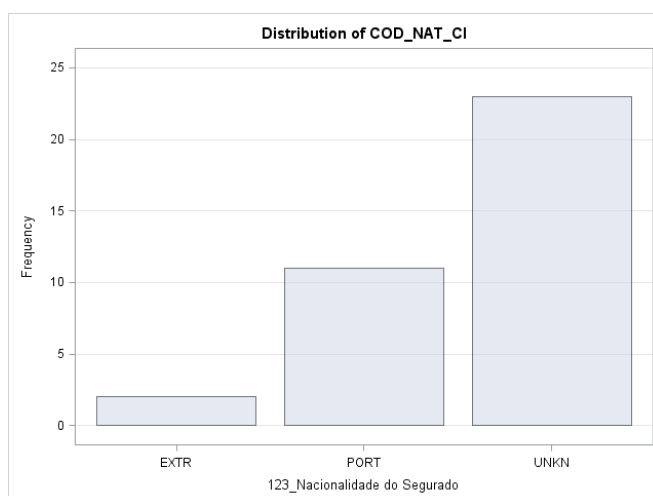


Gráfico 44 – Nacionalidade do sinistrado para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

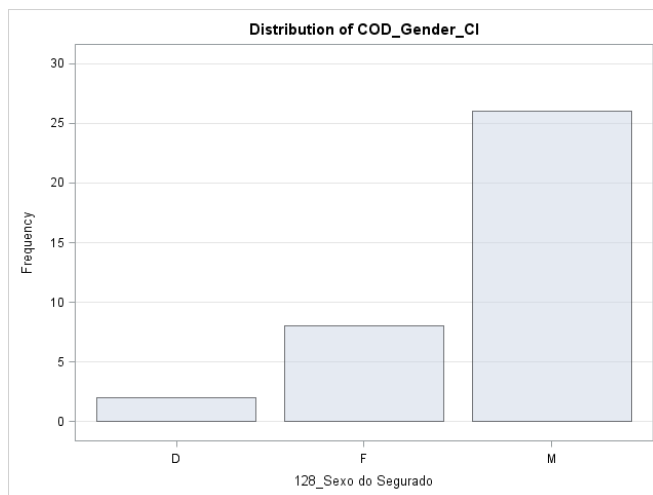


Gráfico 45 – Género do sinistrado para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

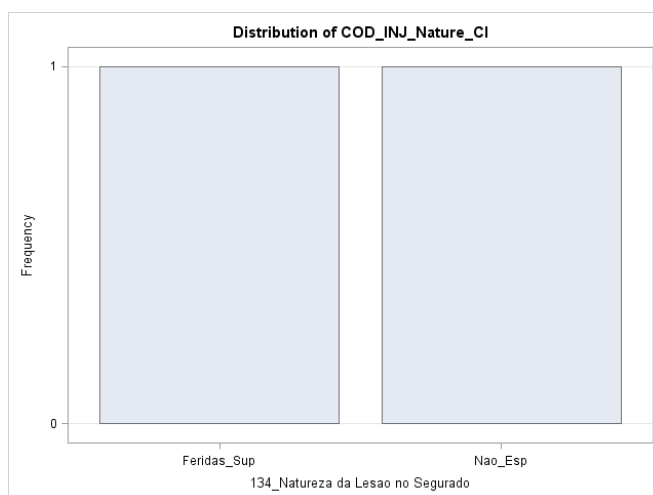


Gráfico 46 – Tipos de lesão decorrente do sinistro para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

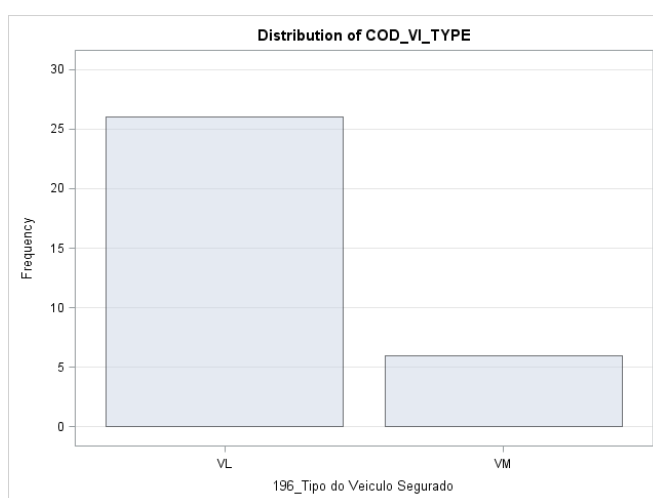


Gráfico 47 – Tipos de veiculo seguro para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

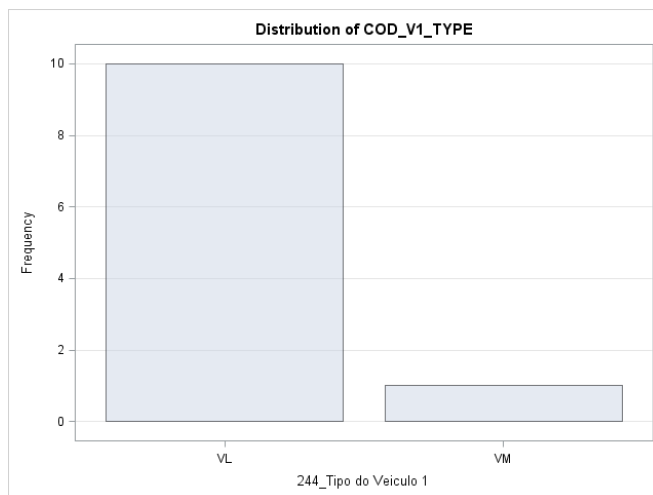


Gráfico 48 – Tipo do veículo 1 no sinistro para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

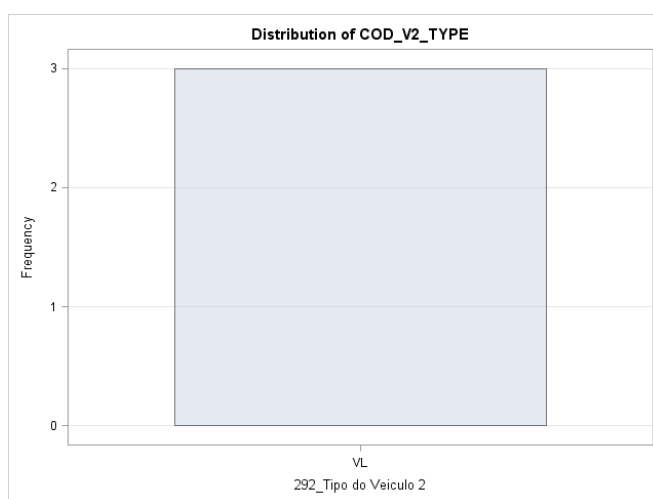


Gráfico 49 – Tipo do veículo 2 no sinistro para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

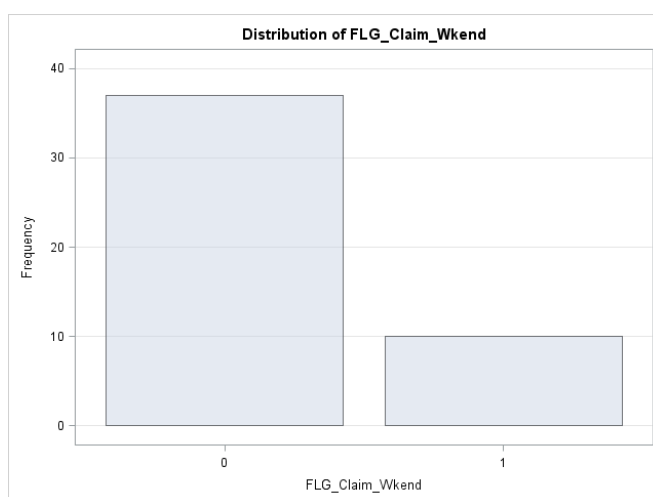


Gráfico 50 – Variável binária indicativa se o acidente ocorreu num dia de semana (0) ou se ocorreu num fim-de-semana/feriado (1) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .



Gráfico 51 – Variável binária indicativa se o acidente ocorreu num em Portugal (0) ou se ocorreu no estrangeiro (1) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

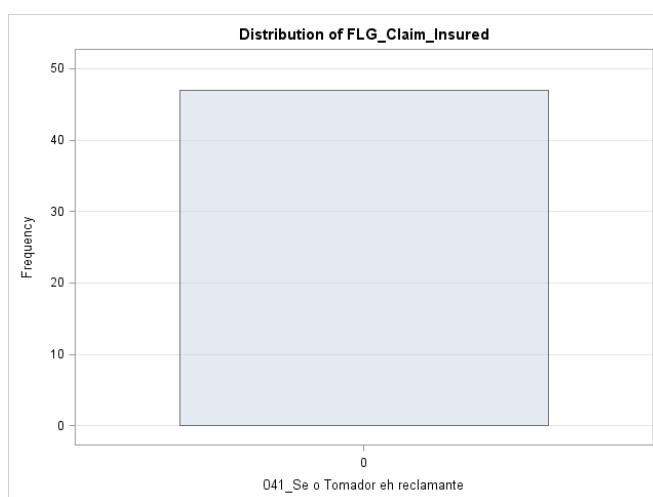


Gráfico 52 – Variável binária indicativa se o tomador da apólice é reclamante (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .



Gráfico 53 – Variável binária indicativa se existiu intervenção das autoridades (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

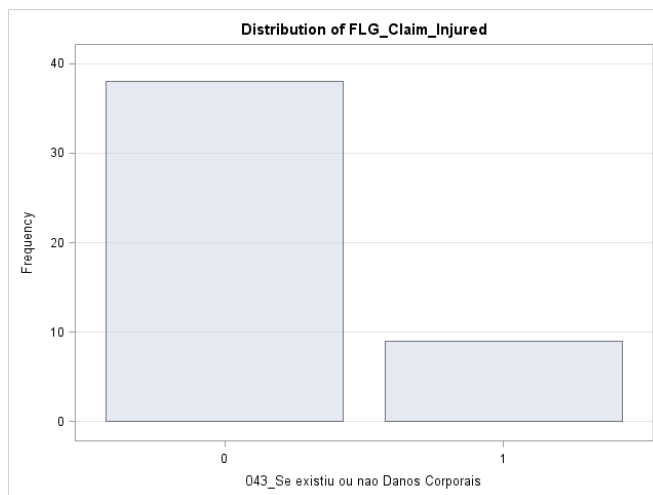


Gráfico 54 – Variável binária indicativa se existiu danos corporais (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .



Gráfico 55 – Variável binária indicativa se existiu testemunhas do sinistro (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

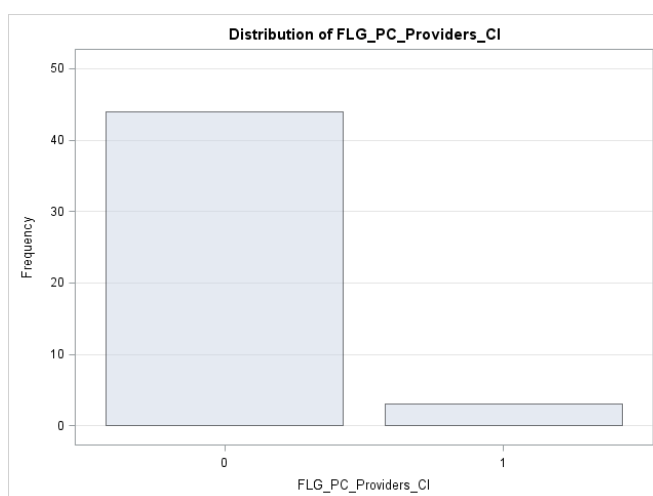


Gráfico 56 – Variável binária que indica se há Sinistros anteriores envolvendo o Segurado e Oficina (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

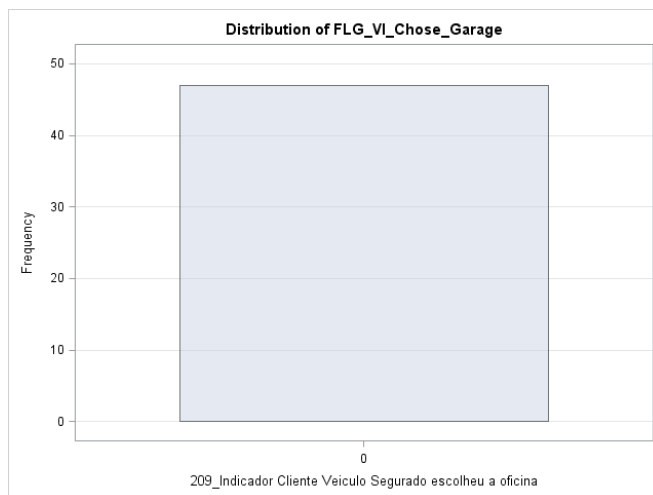


Gráfico 57 – Variável binária que indica se o Segurado escolheu a Oficina (0) ou não (1) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .



Gráfico 58 – Variável binária que indica se a Oficina está associada a fraude (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

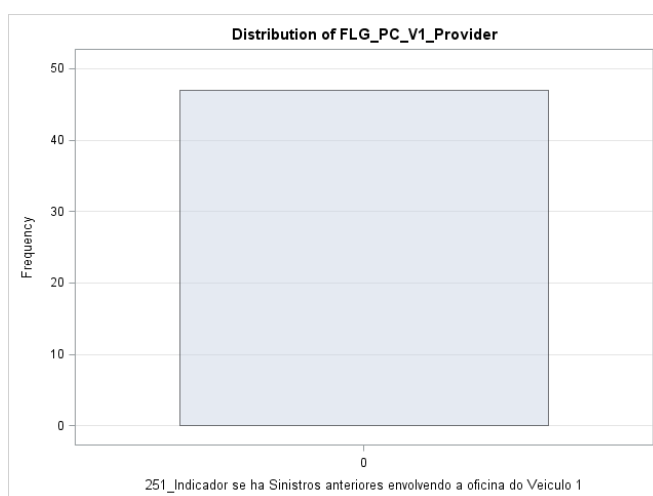


Gráfico 59 – Variável binária se existem sinistros anteriores envolvendo (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .



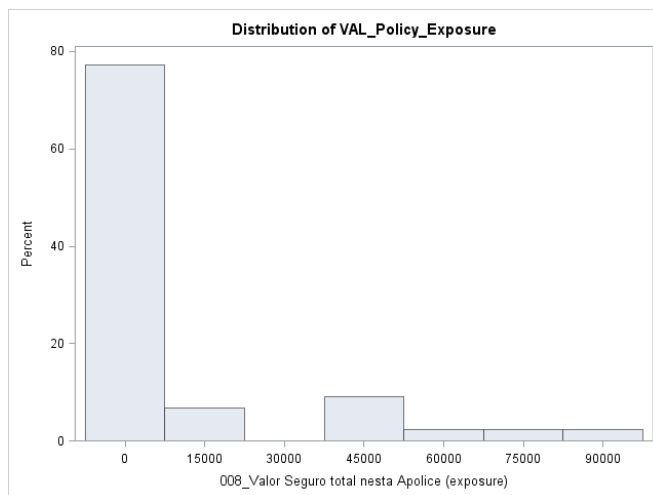


Gráfico 60 – Distribuição do valor seguro pela apólice para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

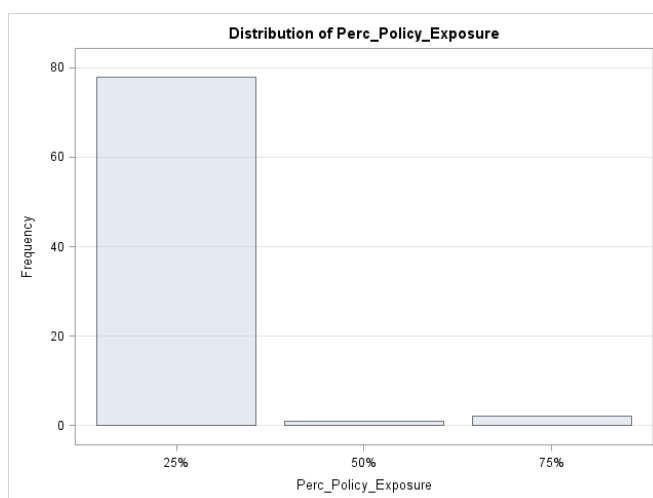


Gráfico 61 – Localização das apólices consoante o quartil de percentagem do valor seguro pela apólice para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

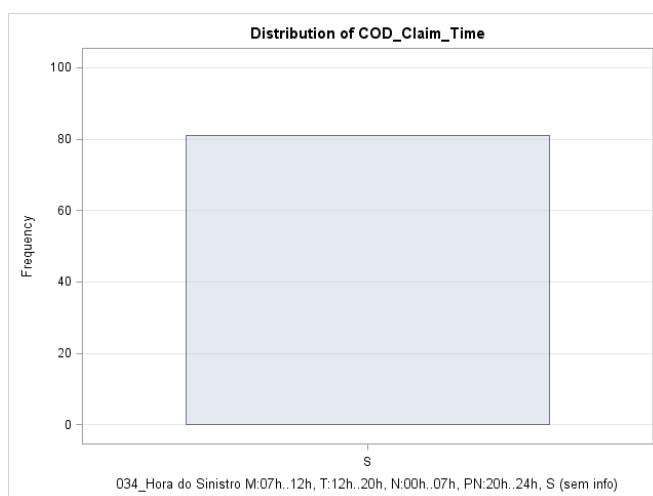


Gráfico 62 – Distribuição temporal da ocorrência dos sinistros para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

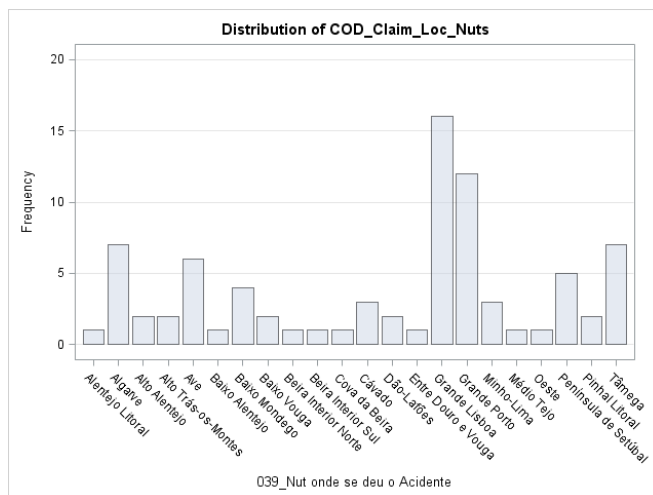


Gráfico 63 – Distribuição geográfica por Nut3 da ocorrência dos sinistros para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

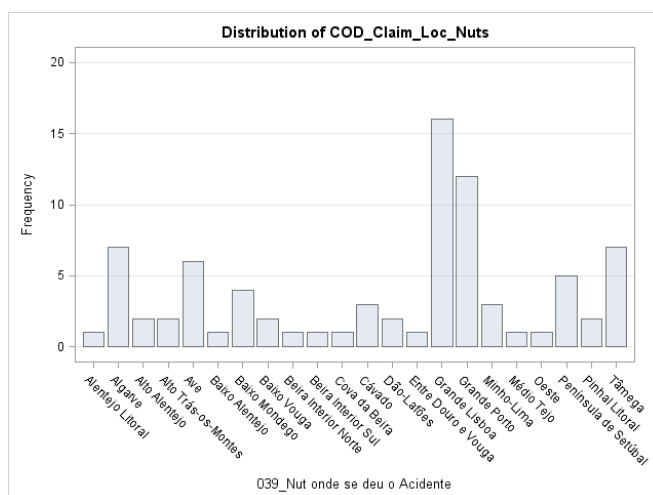


Gráfico 64 – Nacionalidade do sinistrado para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

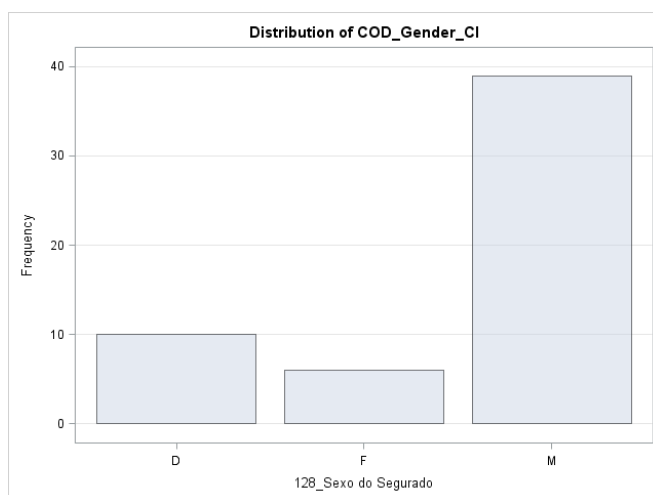


Gráfico 65 – Género do sinistrado para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

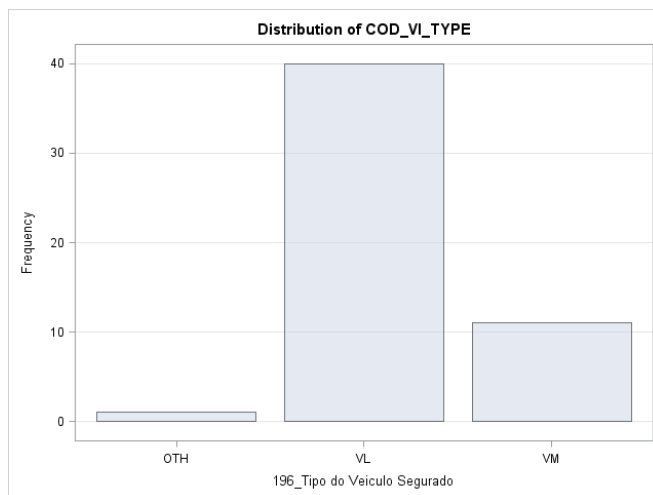


Gráfico 66 – Tipos de veículo seguro para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

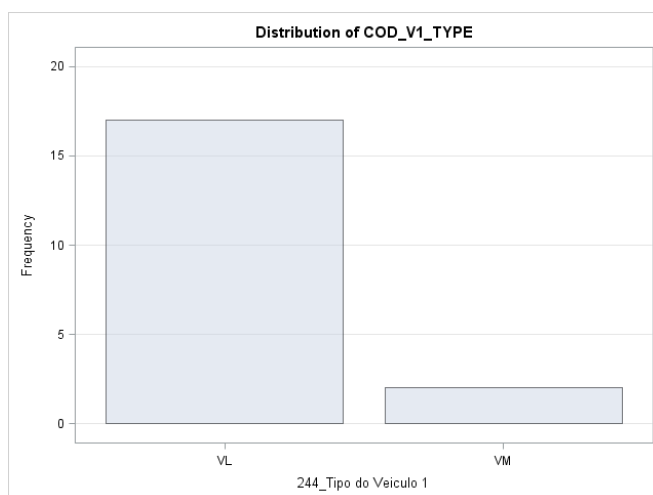


Gráfico 67 – Tipo do veículo 1 no sinistro para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

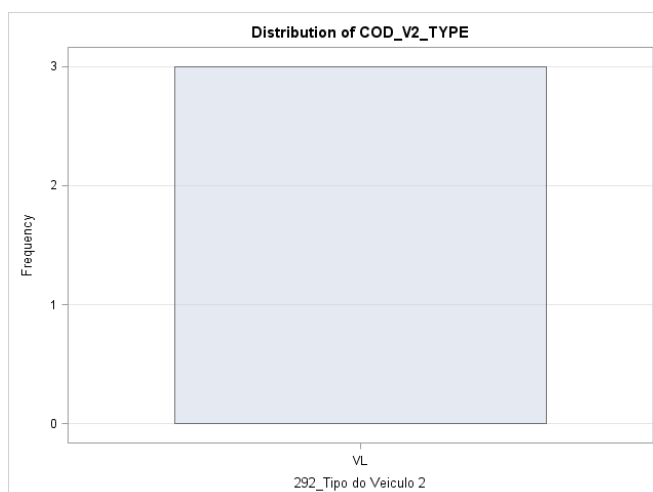


Gráfico 68 – Tipo do veículo 2 no sinistro para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

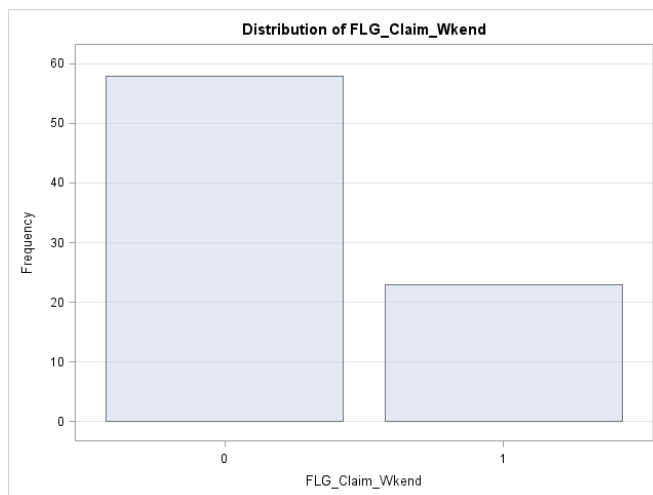


Gráfico 69 – Variável binária indicativa se o acidente ocorreu num dia de semana (0) ou se ocorreu num fim-de-semana/feriado (1) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

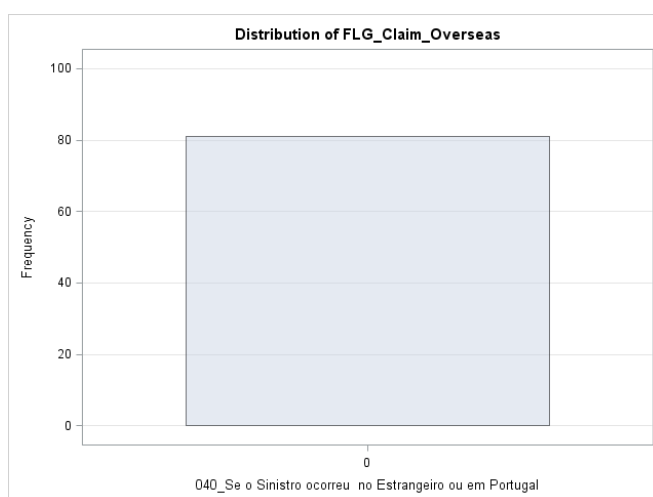


Gráfico 70 – Variável binária indicativa se o acidente ocorreu num em Portugal (0) ou se ocorreu no estrangeiro (1) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

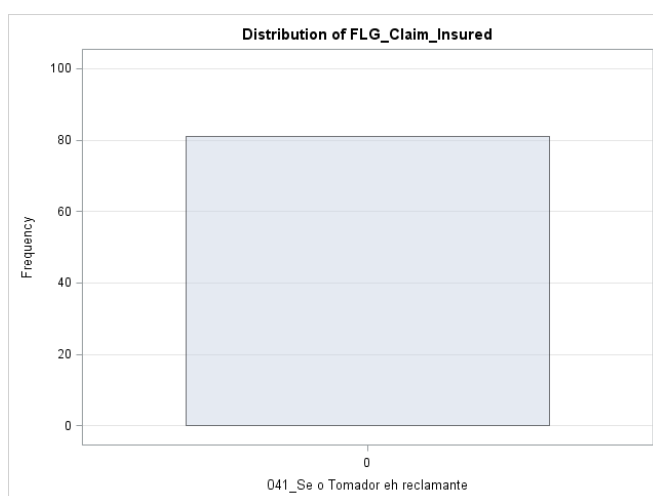


Gráfico 71 – Variável binária indicativa se o tomador da apólice é reclamante (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

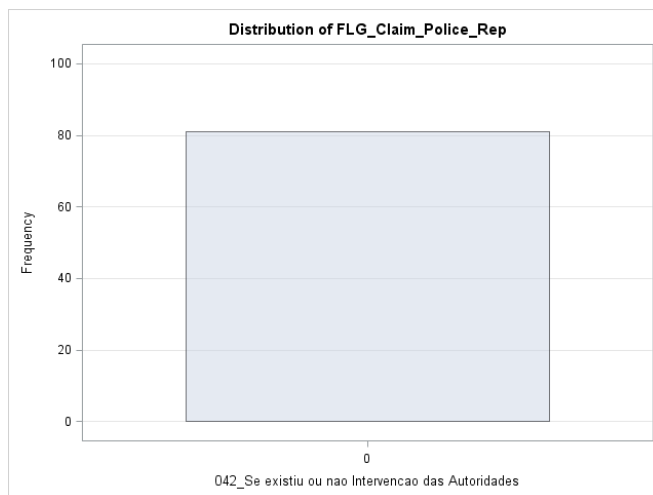


Gráfico 72 – Variável binária indicativa se existiu intervenção das autoridades (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

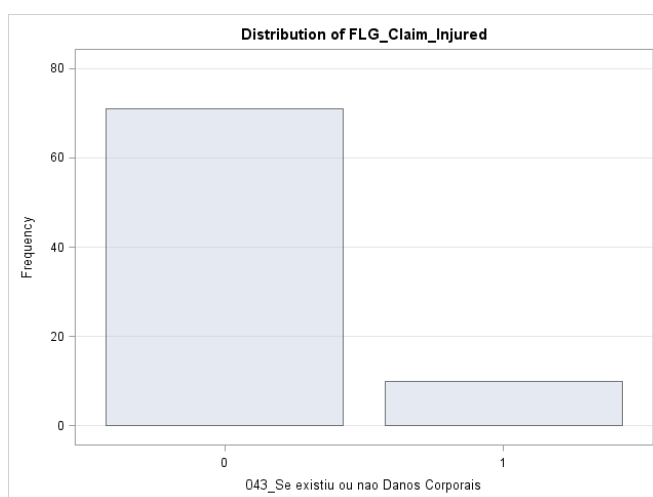


Gráfico 73 – Variável binária indicativa se existiu danos corporais (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

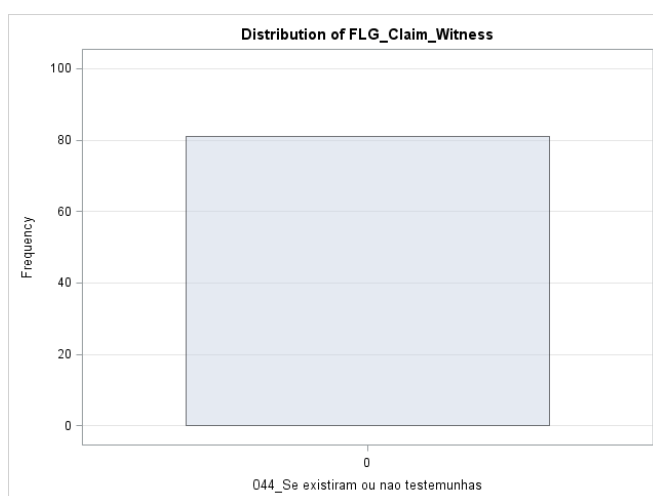


Gráfico 74 – Variável binária indicativa se existiu testemunhas do sinistro (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

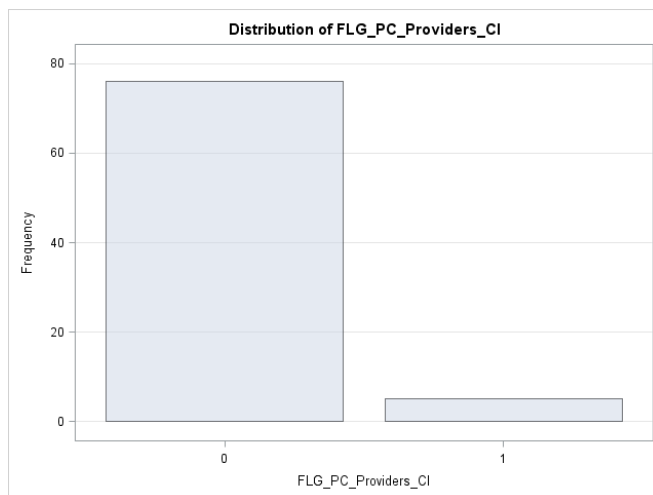


Gráfico 75 – Variável binária que indica se há Sinistros anteriores envolvendo o Segurado e Oficina (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

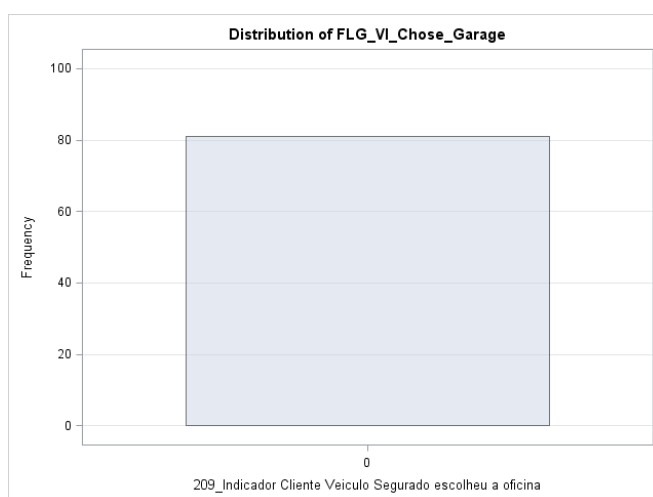


Gráfico 76 – Variável binária que indica se o Segurado escolheu a Oficina (0) ou não (1) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

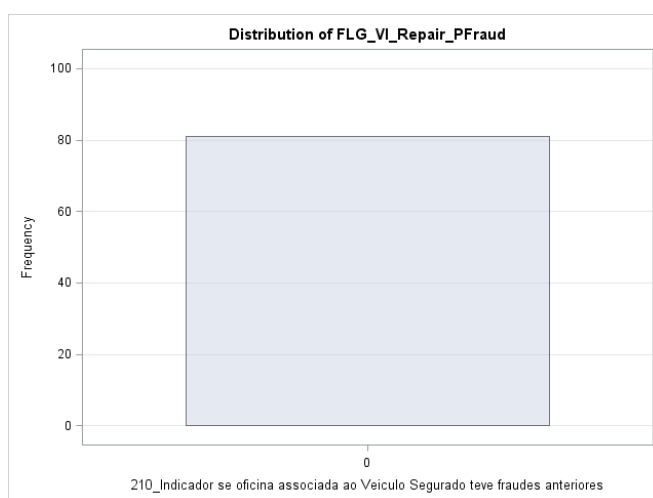


Gráfico 77 – Variável binária que indica se a Oficina está associada a fraude (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

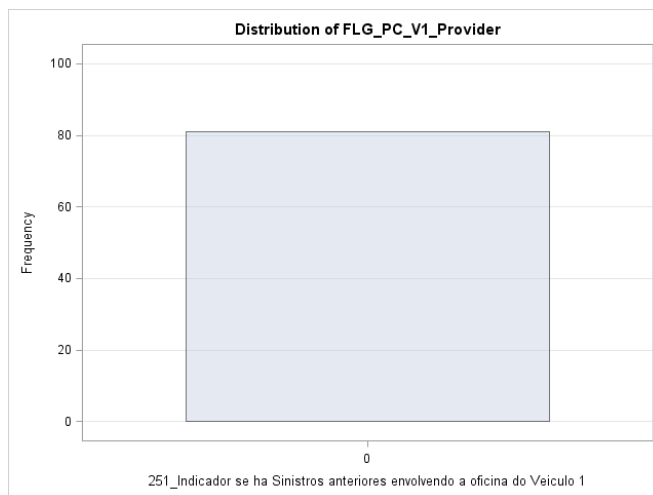


Gráfico 79 – Variável binária se existem sinistros anteriores envolvendo (1) ou não (0) para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

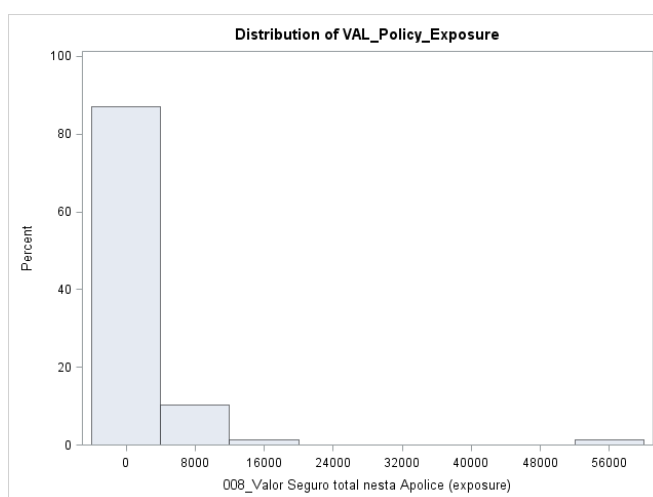


Gráfico 80 – Distribuição do valor seguro pela apólice para os casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 1$ .

## b) Tabelas

Variable	Mean	Std Dev	Minimum	Maximum	N
NUM_Antig_Policy	24.9200534	617.1169	-95891	780	315423
NUM_Antig_Alt_Policy	113.630813	432.801594	-364	7048	333303
VAL_Prop_Exposure	7056.25	29383.67	0	7496136	315423
NUM_Days_Incept	24.6758736	3701.57	-43	2191456	353791
VAL_VI_INSURED_VALUE	2575049.04	10923847.1	0	50000000	73430
NUM_VI_ENGINE	3385.95	12639.64	0	99999	89887
NUM_V1_KM	0	0	0	0	9496
NUM_V1_Age_FR	11.8129169	59.8968346	-1	2011	6147
NUM_V1_ENGINE	1216.41	7456.31	0	99999	5044
NUM_V2_KM	0	0	0	0	981
NUM_V2_ENGINE	1490.66	873.275592	0	5480	410
NUM_V2_Prior_Lines	0.1228388	2.0400152	0	70	4974

Tabela 2 – Estatísticas descritivas para as variáveis numéricas nos casos onde *FLG\_Susp\_Fraud* = 0 e

*FLG\_Conf\_Fraud* = 0.

008_Vvalor Seguro total nesta Apolice (Em valores de percentagem do do valor mais elevado - 25%, 50% e 75%)		
Perc_Policy_Exposure	Frequency	Percent
25%	130879	36.99
50%	164765	46.57
75%	58172	16.44

Tabela 3 – Tabela de frequência para a variável *Perc\_Policy\_Exposure* nos casos onde *FLG\_Susp\_Fraud* = 0 e

*FLG\_Conf\_Fraud* = 0.

034_Hora do Sinistro M:07h..12h, T:12h..20h, N:00h..07h, PN:20h..24h, S (sem info)		
COD_Claim_Time	Frequency	Percent
M	39174	11.07
N	66733	18.86
PN	13379	3.78
S	152380	43.07
T	82150	23.22

Tabela 4 – Tabela de frequência para a variável *COD\_Claim\_Time* nos casos onde *FLG\_Susp\_Fraud* = 0 e

*FLG\_Conf\_Fraud* = 0.

039_Nut onde se deu o Acidente		
COD_Claim_Loc_Nuts	Frequency	Percent
Alentejo Central	2120	0.81
Alentejo Litoral	1033	0.40
Algarve	8957	3.43
Alto Alentejo	1177	0.45
Alto Trás-os-Montes	3407	1.31
Ave	11562	4.43
Açores	3253	1.25
Baixo Alentejo	1271	0.49
Baixo Mondego	6789	2.60
Baixo Vouga	6818	2.61
Beira Interior Norte	1919	0.74
Beira Interior Sul	1672	0.64
Cova da Beira	1331	0.51
Cávado	15475	5.93
Douro	4465	1.71
Dão-Lafões	4541	1.74
Entre Douro e Vouga	4998	1.92
Grande Lisboa	67813	26.01
Grande Porto	53881	20.66
Lezíria do Tejo	4305	1.65
Madeira	3017	1.16
Minho-Lima	4591	1.76
Médio Tejo	3370	1.29
Oeste	5515	2.11
Península de Setúbal	15452	5.93
Pinhal Interior Norte	1087	0.42
Pinhal Interior Sul	331	0.13
Pinhal Litoral	6546	2.51
Serra da Estrela	276	0.11
Tâmega	13790	5.29

Tabela 5 – Tabela de frequência para a variável *COD\_Claim\_Loc\_Nuts* nos casos onde *FLG\_Susp\_Fraud* = 0 e

*FLG\_Conf\_Fraud* = 0.



123_Nacionalidade do Segurado		
COD_NAT_CI	Frequency	Percent
EXTR	815	0.57
PORT	103578	72.98
UNKN	37533	26.45

Tabela 6 – Tabela de frequência para a variável COD\_NAT\_CI nos casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

128_Sexo do Segurado		
COD_Gender_CI	Frequency	Percent
	9047	7.55
D	18785	15.67
F	28002	23.36
M	64051	53.43

Tabela 7 – Tabela de frequência para a variável COD\_Gender\_CI nos casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

COD_Occupation_CI	Frequency	Percent
Profissão detalhada	4058	2.91
NULL	135650	97.09

Tabela 8 – Tabela de frequência para a variável COD\_Occupation\_CI nos casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

134_Natureza da Lesao no Segurado		
COD_INJ_Nature_CI	Frequency	Percent
Choque	30	6.80
Entorses	2	0.45
Feridas_Sup	150	34.01
Fracturas	22	4.99
Lesoes_Int	13	2.95
Multiplas	17	3.85
Nao_Esp	207	46.94

Tabela 9 – Tabela de frequência para a variável COD\_INJ\_Nature\_CI nos casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 0$ .

195_Marca e Modelo Veiculo Segurado		
COD_VI_MM	Frequency	Percent
A.J.P.	2	0.00
AIXAM	134	0.11
ALFA ROMEO	657	0.56
APRILIA	15	0.01
ARO	5	0.00
ASTON MARTIN	3	0.00
AUDI	4999	4.24
AUSTIN	1	0.00
AUTOBIANCHI	1	0.00
BEDFORD	16	0.01
BENALU	1	0.00
BENTLEY	1	0.00
BMW	5634	4.78
CAGIVA	1	0.00
CASAL	3	0.00
CHATENET	9	0.01
CHEVROLET	419	0.36
CHRYSLER	134	0.11
CITROEN	7076	6.00

195_Marca e Modelo Veiculo Segurado		
COD_VI_MM	Frequency	Percent
CPI	3	0.00
DACIA	62	0.05
DAELIM	3	0.00
DAEWOO	177	0.15
DAF	134	0.11
DAIHATSU	40	0.03
DAIMLER BENZ	1	0.00
DATSUN	10	0.01
DAVID BROWN	1	0.00
DODGE	58	0.05
DUCATI	3	0.00
FAMEL	2	0.00
FERRARI	11	0.01
FIAT	4854	4.12
FORD	7292	6.19
GALERA	2	0.00
GALLOPER	4	0.00
GALUCHO	1	0.00
GILERA	9	0.01
HONDA	2598	2.20
HYUNDAI	1066	0.90
INVEPE	10	0.01
ISUZU	105	0.09
IVECO	419	0.36
JAGUAR	167	0.14
JCB	2	0.00
JDM	7	0.01
JEEP	138	0.12
JONWAY	1	0.00
KAWASAKI	25	0.02
KEEWAY	16	0.01
KIA	477	0.40
KRONE	1	0.00
KYMCO	6	0.01
LANCIA	413	0.35
LAND ROVER	354	0.30
LECIBERICA	3	0.00
LECSOR	1	0.00
LEXUS	123	0.10
LIGIER	20	0.02
LOTUS	3	0.00
MACAL	2	0.00
MAN	80	0.07
MASERATI	3	0.00
MASSEY	1	0.00
MASSEY FERGUSON	1	0.00
MAZDA	1082	0.92
MBK	3	0.00
MEGA	4	0.00
MERCEDES	7800	6.62
MG	75	0.06
MICRO	4	0.00
MICROCAR	35	0.03
MINI	328	0.28
MITSUBISHI	2249	1.91
NISSAN	2226	1.89

195_Marca e Modelo Veiculo Segurado		
COD_VI_MM	Frequency	Percent
OPEL	11313	9.60
OUTRO	11554	9.80
Outro	143	0.12
PEGASO	1	0.00
PEUGEOT	8501	7.21
PGO	1	0.00
PIAGGIO	29	0.02
PORSCHE	174	0.15
PORTARO	2	0.00
RENAULT	13202	11.20
ROVER	670	0.57
SAAB	137	0.12
SAMRO	4	0.00
SCANIA	143	0.12
SCHMITZ	7	0.01
SEAT	1763	1.50
SETRA	5	0.00
SIMPA	2	0.00
SIS SACHS	3	0.00
SKODA	915	0.78
SMART	1133	0.96
SSANGYONG	16	0.01
STANDARD	1	0.00
STEYR	7	0.01
SUBARU	46	0.04
SUZUKI	635	0.54
SYM	14	0.01
TALBOT	1	0.00
TATA	8	0.01
TOYOTA	5553	4.71
TRIUMPH	2	0.00
UMM	2	0.00
VESPA	17	0.01
VOLKSWAGEN	8648	7.34
VOLVO	1417	1.20
YAMAHA	185	0.16

Tabela 10 – Tabela de frequência para a variável COD\_VI\_MM nos casos onde *FLG\_Susp\_Fraud* = 0 e

*FLG\_Conf\_Fraud* = 0.

196_Tipo do Veiculo Segurado		
COD_VI_TYPE	Frequency	Percent
OTH	702	0.64
VL	107140	97.36
VM	1691	1.54
VP	514	0.47

Tabela 11 – Tabela de frequência para a variável COD\_VI\_TYPE nos casos onde *FLG\_Susp\_Fraud* = 0 e

*FLG\_Conf\_Fraud* = 0.

243_Marca e Modelo Veiculo 1		
COD_V1_MM	Frequency	Percent
A.J.P.	15	0.04
ADLY	19	0.05

243_Marca e Modelo Veiculo 1		
COD_V1_MM	Frequency	Percent
AIXAM	54	0.14
ALFA ROMEO	254	0.68
APRILIA	13	0.03
ARO	267	0.72
ASIA	2	0.01
AUDI	1332	3.57
AUSTIN	2	0.01
AUTOBIANCHI	1	0.00
BEDFORD	24	0.06
BENALU	1	0.00
BENELLI	1	0.00
BMW	1398	3.75
CASAL	7	0.02
CHATENET	1	0.00
CHEVROLET	111	0.30
CHRYSLER	38	0.10
CITROEN	2145	5.75
DACIA	13	0.03
DAELIM	1	0.00
DAEWOO	66	0.18
DAF	119	0.32
DAIHATSU	22	0.06
DATSUN	1	0.00
DODGE	11	0.03
DUCATI	1	0.00
FAMEL	1	0.00
FIAT	1872	5.02
FORD	2536	6.80
GILERA	9	0.02
HARLEY DAVIDSON	3	0.01
HONDA	1001	2.68
HYUNDAI	376	1.01
INNOCENTI	1	0.00
INVEPE	6	0.02
ISUZU	94	0.25
IVECO	156	0.42
JAGUAR	30	0.08
JEEP	36	0.10
KAWASAKI	28	0.08
KEEWAY	21	0.06
KIA	150	0.40
KTM	3	0.01
LADA	1	0.00
LAMBORGHINI	2	0.01
LANCIA	150	0.40
LAND ROVER	110	0.29
LEXUS	18	0.05
LIGIER	3	0.01
LOTUS	1	0.00
MACAL	1	0.00
MALAGUTI	2	0.01
MAN	120	0.32
MASSEY	1	0.00
MAZDA	294	0.79
MEGA	1	0.00
MERCEDES	2396	6.42

243_Marca e Modelo Veiculo 1		
COD_V1_MM	Frequency	Percent
MG	14	0.04
MICRO	1	0.00
MICROCAR	14	0.04
MINI	64	0.17
MITSUBISHI	848	2.27
MZ	1	0.00
NISSAN	886	2.37
NORTON	1	0.00
OPEL	3627	9.72
OUTRO	1197	3.21
Outro	256	0.69
PEUGEOT	2600	6.97
PGO	1	0.00
PIAGGIO	46	0.12
PORSCHE	47	0.13
RENAULT	4156	11.14
ROVER	286	0.77
SAAB	24	0.06
SACHS	2	0.01
SADO	1	0.00
SANYANG	1	0.00
SCANIA	112	0.30
SCHMITZ	2	0.01
SEAT	997	2.67
SETRA	1	0.00
SKODA	300	0.80
SMART	281	0.75
SSANGYONG	7	0.02
SUBARU	14	0.04
SUZUKI	270	0.72
SYM	9	0.02
TATA	7	0.02
TOYOTA	2025	5.43
TRIUMPH	2	0.01
UMM	1	0.00
VESPA	18	0.05
VOLKSWAGEN	3084	8.27
VOLVO	504	1.35
YAMAHA	257	0.69
ZUNDAP	2	0.01

Tabela 12 – Tabela de frequência para a variável COD\_V1\_MM nos casos onde *FLG\_Susp\_Fraud* = 0 e *FLG\_Conf\_Fraud* = 0.

244_Tipo do Veiculo 1		
COD_V1_TYPE	Frequency	Percent
OTH	2026	4.90
VL	37661	91.01
VM	1082	2.61
VP	612	1.48

Tabela 13 – Tabela de frequência para a variável COD\_V1\_TYPE nos casos onde *FLG\_Susp\_Fraud* = 0 e *FLG\_Conf\_Fraud* = 0

291_Marca e Modelo Veiculo 2		
COD_V2_MM	Frequency	Percent
A.J.P.	14	0.32
ADLY	4	0.09
AIXAM	3	0.07
ALFA ROMEO	41	0.95
APRILIA	2	0.05
ARO	60	1.39
AUDI	156	3.61
BEDFORD	3	0.07
BMW	162	3.75
CASAL	1	0.02
CHEVROLET	11	0.25
CHRYSLER	6	0.14
CITROEN	239	5.53
DACIA	3	0.07
DAEWOO	8	0.19
DAF	17	0.39
DAIHATSU	1	0.02
FIAT	214	4.95
FORD	315	7.29
GILERA	3	0.07
HARLEY DAVIDSON	1	0.02
HONDA	102	2.36
HYUNDAI	40	0.93
ISUZU	5	0.12
IVECO	18	0.42
JAGUAR	3	0.07
JAWA	1	0.02
JEEP	5	0.12
KEEWAY	2	0.05
KIA	17	0.39
LANCIA	14	0.32
LAND ROVER	13	0.30
LEXUS	2	0.05
MAN	6	0.14
MAZDA	40	0.93
MERCEDES	242	5.60
MG	3	0.07
MICROCAR	2	0.05
MINI	7	0.16
MITSUBISHI	95	2.20
NISSAN	83	1.92
NORTON	1	0.02
OPEL	428	9.91
OUTRO	114	2.64
Outro	85	1.97
PEUGEOT	319	7.39
PIAGGIO	2	0.05
PORSCHE	3	0.07
RENAULT	475	11.00
ROVER	33	0.76
SAAB	6	0.14
SACHS	1	0.02
SCANIA	6	0.14
SCHMITZ	1	0.02
SEAT	123	2.85
SETRA	1	0.02

291_Marca e Modelo Veiculo 2		
COD_V2_MM	Frequency	Percent
SKODA	27	0.63
SMART	27	0.63
SSANGYONG	1	0.02
SUBARU	2	0.05
SUZUKI	38	0.88
SYM	3	0.07
TATA	1	0.02
TOYOTA	208	4.82
UMM	1	0.02
VESPA	3	0.07
VOLKSWAGEN	356	8.24
VOLVO	60	1.39
YAMAHA	30	0.69

Tabela 14 – Tabela de frequência para a variável COD\_V2\_MM nos casos onde *FLG\_Susp\_Fraud* = 0 e

*FLG\_Conf\_Fraud* = 0

292_Tipo do Veiculo 2		
COD_V2_TYPE	Frequency	Percent
OTH	89	2.00
VL	4203	94.39
VM	121	2.72
VP	40	0.90

Tabela 15 – Tabela de frequência para a variável COD\_V2\_TYPE nos casos onde *FLG\_Susp\_Fraud* = 0 e

*FLG\_Conf\_Fraud* = 0

035_Se o Sinistro ocorreu ou nao durante um Final de Semana ou Feriado (Claim_dt) (tabela de feriados nacionais)		
FLG_Claim_Wkend	Frequency	Percent
0	276939	78.27
1	76877	21.73

Tabela 16 – Tabela de frequência para a variável FLG\_Claim\_Wkend nos casos onde *FLG\_Susp\_Fraud* = 0 e

*FLG\_Conf\_Fraud* = 0

040_Se o Sinistro ocorreu no Estrangeiro ou em Portugal		
FLG_Claim_Overseas	Frequency	Percent
0	351196	99.26
1	2620	0.74

Tabela 17 – Tabela de frequência para a variável FLG\_Claim\_Overseas nos casos onde *FLG\_Susp\_Fraud* = 0 e

*FLG\_Conf\_Fraud* = 0

041_Se o Tomador eh reclamante		
FLG_Claim_Insured	Frequency	Percent
0	328893	92.96
1	24923	7.04

Tabela 18 – Tabela de frequência para a variável FLG\_Claim\_Insured nos casos onde FLG\_Susp\_Fraud = 0 e  
FLG\_Conf\_Fraud = 0

<b>042_Se existiu ou nao Intervencao das Autoridades</b>		
<b>FLG_Claim_Police_Rep</b>	<b>Frequency</b>	<b>Percent</b>
0	351046	99.22
1	2770	0.78

Tabela 19 – Tabela de frequência para a variável FLG\_Claim\_Police\_Rep nos casos onde FLG\_Susp\_Fraud = 0  
e FLG\_Conf\_Fraud = 0

<b>043_Se existiu ou nao Danos Corporais</b>		
<b>FLG_Claim_Injured</b>	<b>Frequency</b>	<b>Percent</b>
0	341768	96.59
1	12048	3.41

Tabela 20 – Tabela de frequência para a variável FLG\_Claim\_Injured nos casos onde FLG\_Susp\_Fraud = 0 e  
FLG\_Conf\_Fraud = 0

<b>044_Se existiram ou nao testemunhas</b>		
<b>FLG_Claim_Witness</b>	<b>Frequency</b>	<b>Percent</b>
0	353444	99.89
1	372	0.11

Tabela 21 – Tabela de frequência para a variável FLG\_Claim\_Witness nos casos onde FLG\_Susp\_Fraud = 0 e  
FLG\_Conf\_Fraud = 0

<b>138_Indicador se ha Sinistros anteriores envolvendo o Segurado e Oficina em que seu veiculo foi reparado</b>		
<b>FLG_PC_Providers_CI</b>	<b>Frequency</b>	<b>Percent</b>
0	341078	96.40
1	12738	3.60

Tabela 22 – Tabela de frequência para a variável FLG\_PC\_Providers\_CI nos casos onde FLG\_Susp\_Fraud = 0 e  
FLG\_Conf\_Fraud = 0

<b>209_Indicador Cliente Veiculo Segurado escolheu a oficina</b>		
<b>FLG_VI_Chose_Garage</b>	<b>Frequency</b>	<b>Percent</b>
0	353816	100.00

Tabela 23 – Tabela de frequência para a variável FLG\_VI\_Chose\_Garage nos casos onde FLG\_Susp\_Fraud = 0 e  
FLG\_Conf\_Fraud = 0

<b>210_Indicador se oficina associada ao Veiculo Segurado teve fraudes anteriores</b>		
<b>FLG_VI_Repair_PFraud</b>	<b>Frequency</b>	<b>Percent</b>
0	346384	97.90
1	7432	2.10

Tabela 24 – Tabela de frequência para a variável FLG\_VI\_Repair\_PFraud nos casos onde FLG\_Susp\_Fraud = 0  
e FLG\_Conf\_Fraud = 0



251_Indicador se ha Sinistros anteriores envolvendo a oficina do Veiculo 1		
FLG_PC_V1_Provider	Frequency	Percent
0	351698	99.40
1	2118	0.60

Tabela 25 – Tabela de frequência para a variável FLG\_PC\_V1\_Provider nos casos onde FLG\_Susp\_Fraud = 0 e FLG\_Conf\_Fraud = 0

Analysis Variable : VAL_Policy_Exposure 008_Valor Seguro total nesta Apolice (exposure)				
Mean	Std Dev	Minimum	Maximum	N
13959.29	16005.16	0	99984.00	311256

Tabela 26 – Tabela de estatísticas descritivas para a variável VAL\_Policy\_Exposure nos casos onde FLG\_Susp\_Fraud = 0 e FLG\_Conf\_Fraud = 0

Variable	Mean	Std Dev	Minimum	Maximum	N
NUM_Antig_Policy	18.0465574	27.7481845	0	235	1525
NUM_Antig_Alt_Policy	22.6847969	147.8461216	-356	364	1526
VAL_Prop_Exposure	3607.68	10441.52	0	212553.68	1525
NUM_Days_Incept	14.0597791	23.6081539	0	220	1539
VAL_VI_INSURED_VALUE	44.8772689	394.2821398	0	6000	238
NUM_VI_ENGINE	2458.43	7657.76	0	99999	334
NUM_V1_KM	.	.	.	.	0
NUM_V1_Age_FR	7.25	5.8002463	0	19	8
NUM_V1_ENGINE	1667.33	471.4332756	1124	1968	3
NUM_V2_KM	.	.	.	.	0
NUM_V2_ENGINE	1968	.	1968	1968	1
NUM_V2_Prior_Lines	0.025641	0.1601282	0	1	39

Tabela 27 – Estatísticas descritivas para as variáveis numéricas nos casos onde FLG\_Susp\_Fraud = 0 e FLG\_Conf\_Fraud = 1.

008_Valor Seguro total nesta Apolice (Em valores de percentagem do do valor mais elevado - 25%, 50% e 75%)		
Perc_Policy_Exposure	Frequency	Percent
25%	474	30.80
50%	901	58.54
75%	164	10.66

Tabela 28 – Tabela de frequência para a variável Perc\_Policy\_Exposure nos casos onde FLG\_Susp\_Fraud = 0 e FLG\_Conf\_Fraud = 1.

034_Hora do Sinistro M:07h..12h, T:12h..20h, N:00h..07h, PN:20h..24h, S (sem info)		
COD_Claim_Time	Frequency	Percent
M	168	10.92
N	663	43.08
PN	108	7.02
S	309	20.08
T	291	18.91

Tabela 29 – Tabela de frequência para a variável COD\_Claim\_Time nos casos onde FLG\_Susp\_Fraud = 0 e FLG\_Conf\_Fraud = 1.

039_Nut onde se deu o Acidente
--------------------------------

COD_Claim_Loc_Nuts	Frequency	Percent
Alentejo Central	4	0.37
Alentejo Litoral	2	0.19
Algarve	47	4.39
Alto Alentejo	4	0.37
Alto Trás-os-Montes	12	1.12
Ave	73	6.82
Açores	15	1.40
Baixo Alentejo	5	0.47
Baixo Mondego	32	2.99
Baixo Vouga	25	2.34
Beira Interior Norte	4	0.37
Beira Interior Sul	9	0.84
Cova da Beira	3	0.28
Cávado	90	8.41
Douro	29	2.71
Dão-Lafões	21	1.96
Entre Douro e Vouga	27	2.52
Grande Lisboa	209	19.53
Grande Porto	197	18.41
Lezíria do Tejo	29	2.71
Madeira	15	1.40
Minho-Lima	16	1.50
Médio Tejo	10	0.93
Oeste	23	2.15
Península de Setúbal	64	5.98
Pinhal Interior Norte	5	0.47
Pinhal Interior Sul	2	0.19
Pinhal Litoral	16	1.50
Tâmega	82	7.66

Tabela 30 – Tabela de frequência para a variável COD\_Claim\_Loc\_Nuts nos casos onde

*FLG\_Susp\_Fraud* = 0 e *FLG\_Conf\_Fraud* = 1.

123_Nacionalidade do Segurado		
COD_NAT_CI	Frequency	Percent
EXTR	7	1.52
PORT	365	79.18
UNKN	89	19.31

Tabela 31 – Tabela de frequência para a variável COD\_NAT\_CI nos casos onde *FLG\_Susp\_Fraud* = 0 e

*FLG\_Conf\_Fraud* = 1.

128_Sexo do Segurado		
COD_Gender_CI	Frequency	Percent
D	34	9.02
F	79	20.95
M	264	70.03

Tabela 32 – Tabela de frequência para a variável COD\_Gender\_CI nos casos onde *FLG\_Susp\_Fraud* = 0

e *FLG\_Conf\_Fraud* = 1.

COD_Occupation_CI	Frequency	Percent
Profissão detalhada	14	3.05
NULL	445	96.95

Tabela 33 – Tabela de frequência para a variável COD\_Occupation\_CI nos casos onde

$FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$ .

134_Natureza da Lesao no Segurado		
COD_INJ_Nature_CI	Frequency	Percent
Choque	1	33.33
Feridas_Sup	2	66.67

Tabela 34 – Tabela de frequência para a variável COD\_INJ\_Nature\_CI nos casos onde  $FLG\_Susp\_Fraud$

$= 0$  e  $FLG\_Conf\_Fraud = 1$ .

195_Marca e Modelo Veiculo Segurado		
COD_VI_MM	Frequency	Percent
AIXAM	1	0.23
ALFA ROMEO	2	0.46
AUDI	42	9.72
BMW	27	6.25
CHEVROLET	1	0.23
CITROEN	35	8.10
DAF	1	0.23
FIAT	18	4.17
FORD	32	7.41
HONDA	15	3.47
HYUNDAI	3	0.69
IVECO	2	0.46
KAWASAKI	3	0.69
KIA	1	0.23
LAND ROVER	1	0.23
MASERATI	1	0.23
MAZDA	2	0.46
MERCEDES	43	9.95
MG	2	0.46
MINI	2	0.46
MITSUBISHI	12	2.78
NISSAN	10	2.31
OPEL	35	8.10
OUTRO	3	0.69
Outro	2	0.46
PEUGEOT	22	5.09
PIAGGIO	1	0.23
PORSCHE	1	0.23
RENAULT	37	8.56
ROVER	3	0.69
SAMRO	1	0.23
SCANIA	1	0.23
SEAT	4	0.93
SKODA	1	0.23
SMART	3	0.69
SUBARU	1	0.23
SUZUKI	4	0.93
TOYOTA	19	4.40
VESPA	1	0.23
VOLKSWAGEN	28	6.48
VOLVO	5	1.16
YAMAHA	4	0.93

Tabela 35 – Tabela de frequência para a variável COD\_VI\_MM nos casos onde  $FLG\_Susp\_Fraud = 0$  e

$FLG\_Conf\_Fraud = 1$ .

196_Tipo do Veiculo Segurado		
COD_VI_TYPE	Frequency	Percent
OTH	3	0.67
VL	419	94.16
VM	20	4.49
VP	3	0.67

Tabela 36 – Tabela de frequência para a variável COD\_VI\_TYPE nos casos onde *FLG\_Susp\_Fraud* = 0 e *FLG\_Conf\_Fraud* = 1.

243_Marca e Modelo Veiculo 1		
COD_V1_MM	Frequency	Percent
AIXAM	2	0.76
ALFA ROMEO	2	0.76
ARO	3	1.15
AUDI	19	7.25
BMW	14	5.34
CHEVROLET	1	0.38
CITROEN	16	6.11
DACIA	1	0.38
DAEWOO	1	0.38
DAF	1	0.38
FIAT	11	4.20
FORD	10	3.82
HONDA	10	3.82
HYUNDAI	2	0.76
KAWASAKI	1	0.38
KEEWAY	1	0.38
MAZDA	1	0.38
MERCEDES	19	7.25
MITSUBISHI	6	2.29
NISSAN	8	3.05
OPEL	31	11.83
OUTRO	7	2.67
Outro	1	0.38
PEUGEOT	24	9.16
PORSCHE	2	0.76
RENAULT	16	6.11
ROVER	1	0.38
SCANIA	1	0.38
SEAT	3	1.15
SKODA	4	1.53
SMART	4	1.53
SUZUKI	2	0.76
TOYOTA	8	3.05
VOLKSWAGEN	20	7.63
VOLVO	5	1.91
YAMAHA	4	1.53

Tabela 37 – Tabela de frequência para a variável COD\_V1\_MM nos casos onde *FLG\_Susp\_Fraud* = 0 e *FLG\_Conf\_Fraud* = 1.

244_Tipo do Veiculo 1		
COD_V1_TYPE	Frequency	Percent
OTH	1	0.34
VL	270	93.10
VM	17	5.86

244_Tipo do Veiculo 1		
COD_V1_TYPE	Frequency	Percent
VP	2	0.69

Tabela 38 – Tabela de frequência para a variável COD\_V1\_TYPE nos casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$

291_Marca e Modelo Veiculo 2		
COD_V2_MM	Frequency	Percent
ARO	1	3.23
AUDI	3	9.68
BMW	1	3.23
CITROEN	2	6.45
FIAT	3	9.68
FORD	3	9.68
JEEP	1	3.23
KIA	1	3.23
MAZDA	1	3.23
MERCEDES	2	6.45
NISSAN	1	3.23
OPEL	3	9.68
PEUGEOT	3	9.68
RENAULT	1	3.23
SEAT	1	3.23
SUZUKI	2	6.45
VOLKSWAGEN	2	6.45

Tabela 39 – Tabela de frequência para a variável COD\_V2\_MM nos casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$

292_Tipo do Veiculo 2		
COD_V2_TYPE	Frequency	Percent
VL	32	94.12
VM	2	5.88

Tabela 40 – Tabela de frequência para a variável COD\_V2\_TYPE nos casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$

035_Se o Sinistro ocorreu ou nao durante um Final de Semana ou Feriado (Claim_dt) (tabela de feriados nacionais)		
FLG_Claim_Wkend	Frequency	Percent
0	1151	74.79
1	388	25.21

Tabela 41 – Tabela de frequência para a variável FLG\_Claim\_Wkend nos casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$

040_Se o Sinistro ocorreu no Estrangeiro ou em Portugal		
FLG_Claim_Overseas	Frequency	Percent
0	1534	99.68
1	5	0.32

Tabela 42 – Tabela de frequência para a variável FLG\_Claim\_Overseas nos casos onde  $FLG\_Susp\_Fraud = 0$  e  $FLG\_Conf\_Fraud = 1$

041_Se o Tomador eh reclamante		
FLG_Claim_Insured	Frequency	Percent
0	1539	100.00

Tabela 43 – Tabela de frequência para a variável FLG\_Claim\_Insured nos casos onde *FLG\_Susp\_Fraud* = 0 e *FLG\_Conf\_Fraud* = 1

042_Se existiu ou nao Intervencao das Autoridades		
FLG_Claim_Police_Rep	Frequency	Percent
0	1539	100.00

Tabela 44 – Tabela de frequência para a variável FLG\_Claim\_Police\_Rep nos casos onde *FLG\_Susp\_Fraud* = 0 e *FLG\_Conf\_Fraud* = 1

043_Se existiu ou nao Danos Corporais		
FLG_Claim_Injured	Frequency	Percent
0	1365	88.69
1	174	11.31

Tabela 45 – Tabela de frequência para a variável FLG\_Claim\_Injured nos casos onde *FLG\_Susp\_Fraud* = 0 e *FLG\_Conf\_Fraud* = 1

044_Se existiram ou nao testemunhas		
FLG_Claim_Witness	Frequency	Percent
0	1539	100.00

Tabela 46 – Tabela de frequência para a variável FLG\_Claim\_Witness nos casos onde *FLG\_Susp\_Fraud* = 0 e *FLG\_Conf\_Fraud* = 1

138_Indicador se ha Sinistros anteriores envolvendo o Segurado e Oficina em que seu veiculo foi reparado		
FLG_PC_Providers_CI	Frequency	Percent
0	1500	97.47
1	39	2.53

Tabela 47 – Tabela de frequência para a variável FLG\_PC\_Providers\_CI nos casos onde *FLG\_Susp\_Fraud* = 0 e *FLG\_Conf\_Fraud* = 1

209_Indicador Cliente Veiculo Segurado escolheu a oficina		
FLG_VI_Chose_Garage	Frequency	Percent
0	1539	100.00

Tabela 48 – Tabela de frequência para a variável FLG\_VI\_Chose\_Garage nos casos onde *FLG\_Susp\_Fraud* = 0 e *FLG\_Conf\_Fraud* = 1

210_Indicador se oficina associada ao Veiculo Segurado teve fraudes anteriores		
FLG_VI_Repair_PFraud	Frequency	Percent
0	1539	100.00

Tabela 49 – Tabela de frequência para a variável FLG\_VI\_Repair\_PFraud nos casos onde *FLG\_Susp\_Fraud* = 0 e *FLG\_Conf\_Fraud* = 1

251_Indicador se ha Sinistros anteriores envolvendo a oficina do Veiculo 1		
FLG_PC_V1_Provider	Frequency	Percent
0	1539	100.00

Tabela 50 – Tabela de frequência para a variável FLG\_PC\_V1\_Provider nos casos onde *FLG\_Susp\_Fraud* = 0 e *FLG\_Conf\_Fraud* = 1

Analysis Variable : VAL_Policy_Exposure 008_Valor Seguro total nesta Apolice (exposure)				
Mean	Std Dev	Minimum	Maximum	N
10793.62	11514.22	0	85000.00	1521

Tabela 51 – Tabela de estatísticas descritivas para a variável VAL\_Policy\_Exposure nos casos onde

*FLG\_Susp\_Fraud* = 0 e *FLG\_Conf\_Fraud* = 1

Variable	Mean	Std Dev	Minimum	Maximum	N
NUM_Antig_Policy	12.77273	21.4355319	0	130	44
NUM_Antig_Alt_Policy	161.1818	128.1282556	0	360	44
VAL_Prop_Exposure	3233.8	7159.21	0	34018.02	44
NUM_Days_Incept	4.531915	4.8582116	0	21	47
VAL_VI_INSURED_VALUE	.	.	.	.	0
NUM_VI_ENGINE	.	.	.	.	0
NUM_V1_KM	.	.	.	.	0
NUM_V1_Age_FR	4	1.4142136	3	5	2
NUM_V1_ENGINE	.	.	.	.	0
NUM_V2_KM	.	.	.	.	0
NUM_V2_ENGINE	.	.	.	.	0
NUM_V2_Prior_Lines	0	0	0	0	3

Tabela 52 – Estatísticas descritivas para as variáveis numéricas nos casos onde *FLG\_Susp\_Fraud* = 1 e

*FLG\_Conf\_Fraud* = 0.

008_Valor Seguro total nesta Apolice (Em valores de percentagem do do valor mais elevado - 25%, 50% e 75%)		
Perc_Policy_Exposure	Frequency	Percent
25%	37	78.72
50%	3	6.38
75%	7	14.89

Tabela 53 – Tabela de frequência para a variável Perc\_Policy\_Exposure nos casos onde *FLG\_Susp\_Fraud* = 1 e

*FLG\_Conf\_Fraud* = 0.

034_Hora do Sinistro M:07h..12h, T:12h..20h, N:00h..07h, PN:20h..24h, S (sem info)		
COD_Claim_Time	Frequency	Percent
S	47	100.00

Tabela 54 – Tabela de frequência para a variável COD\_Claim\_Time nos casos onde *FLG\_Susp\_Fraud* =

1 e *FLG\_Conf\_Fraud* = 0.

039_Nut onde se deu o Acidente		
COD_Claim_Loc_Nuts	Frequency	Percent
Alentejo Litoral	2	4.26
Algarve	3	6.38
Alto Alentejo	1	2.13
Ave	4	8.51
Baixo Mondego	1	2.13
Baixo Vouga	1	2.13
Beira Interior Norte	1	2.13
Beira Interior Sul	1	2.13
Cávado	2	4.26
Douro	3	6.38
Entre Douro e Vouga	1	2.13
Grande Lisboa	17	36.17
Grande Porto	5	10.64
Médio Tejo	1	2.13

039_Nut onde se deu o Acidente		
COD_Claim_Loc_Nuts	Frequency	Percent
Oeste	1	2.13
Península de Setúbal	2	4.26
Tâmega	1	2.13

Tabela 55 – Tabela de frequência para a variável COD\_Claim\_Loc\_Nuts nos casos onde  
*FLG\_Susp\_Fraud = 1 e FLG\_Conf\_Fraud = 0.*

123_Nacionalidade do Segurado		
COD_NAT_CI	Frequency	Percent
EXTR	2	5.56
PORT	11	30.56
UNKN	23	63.89

Tabela 56 – Tabela de frequência para a variável COD\_NAT\_CI nos casos onde *FLG\_Susp\_Fraud = 1 e*  
*FLG\_Conf\_Fraud = 0.*

128_Sexo do Segurado		
COD_Gender_CI	Frequency	Percent
D	2	5.56
F	8	22.22
M	26	72.22

Tabela 57 – Tabela de frequência para a variável COD\_Gender\_CI nos casos onde *FLG\_Susp\_Fraud = 1*  
*e FLG\_Conf\_Fraud = 0.*

COD_Occupation_CI	Frequency	Percent
Profissão detalhada	1	2.78
NULL	35	97.22

Tabela 58 – Tabela de frequência para a variável COD\_Occupation\_CI nos casos onde  
*FLG\_Susp\_Fraud = 1 e FLG\_Conf\_Fraud = 0.*

134_Natureza da Lesao no Segurado		
COD_INJ_Nature_CI	Frequency	Percent
Feridas_Sup	1	50.00
Nao_Esp	1	50.00

Tabela 59 – Tabela de frequência para a variável COD\_INJ\_Nature\_CI nos casos onde *FLG\_Susp\_Fraud*  
*= 1 e FLG\_Conf\_Fraud = 0.*

195_Marca e Modelo Veiculo Segurado		
COD_VI_MM	Frequency	Percent
ARO	1	3.13
AUDI	1	3.13
BMW	1	3.13
CHEVROLET	1	3.13
CITROEN	2	6.25
DAF	1	3.13
FIAT	1	3.13
FORD	3	9.38
HONDA	2	6.25
KAWASAKI	1	3.13
LAND ROVER	1	3.13
MERCEDES	4	12.50
NISSAN	2	6.25
OPEL	2	6.25
RENAULT	1	3.13
ROVER	1	3.13



195_Marca e Modelo Veiculo Segurado		
COD_VI_MM	Frequency	Percent
SUZUKI	1	3.13
TOYOTA	3	9.38
VOLKSWAGEN	2	6.25
YAMAHA	1	3.13

Tabela 60 – Tabela de frequência para a variável COD\_VI\_MM nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 0.

196_Tipo do Veiculo Segurado		
COD_VI_TYPE	Frequency	Percent
VL	26	81.25
VM	6	18.75

Tabela 61 – Tabela de frequência para a variável COD\_VI\_TYPE nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 0.

243_Marca e Modelo Veiculo 1		
COD_V1_MM	Frequency	Percent
ALFA ROMEO	2	18.18
ARO	1	9.09
FIAT	1	9.09
FORD	1	9.09
OPEL	1	9.09
PEUGEOT	1	9.09
RENAULT	2	18.18
SUZUKI	1	9.09
TOYOTA	1	9.09

Tabela 62 – Tabela de frequência para a variável COD\_V1\_MM nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 0.

244_Tipo do Veiculo 1		
COD_V1_TYPE	Frequency	Percent
VL	10	90.91
VM	1	9.09

Tabela 63 – Tabela de frequência para a variável COD\_V1\_TYPE nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 0.

291_Marca e Modelo Veiculo 2		
COD_V2_MM	Frequency	Percent
ALFA ROMEO	1	33.33
LAND ROVER	1	33.33
OPEL	1	33.33

Tabela 64 – Tabela de frequência para a variável COD\_V2\_MM nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 0.

292_Tipo do Veiculo 2		
COD_V2_TYPE	Frequency	Percent
VL	3	100.00

Tabela 65 – Tabela de frequência para a variável COD\_V2\_TYPE nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 0.

**035\_Se o Sinistro ocorreu ou nao durante um Final de Semana ou Feriado (Claim\_dt) (tabela de feriados nacionais)**

FLG_Claim_Wkend	Frequency	Percent
0	37	78.72
1	10	21.28

Tabela 66 – Tabela de frequência para a variável FLG\_Claim\_Wkend nos casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

040_Se o Sinistro ocorreu no Estrangeiro ou em Portugal		
FLG_Claim_Overseas	Frequency	Percent
0	47	100.00

Tabela 67 – Tabela de frequência para a variável FLG\_Claim\_Overseas nos casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

041_Se o Tomador eh reclamante		
FLG_Claim_Insured	Frequency	Percent
0	47	100.00

Tabela 68 – Tabela de frequência para a variável FLG\_Claim\_Insured nos casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

042_Se existiu ou nao Intervencao das Autoridades		
FLG_Claim_Police_Rep	Frequency	Percent
0	47	100.00

Tabela 69 – Tabela de frequência para a variável FLG\_Claim\_Police\_Rep nos casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

043_Se existiu ou nao Danos Corporais		
FLG_Claim_Injured	Frequency	Percent
0	38	80.85
1	9	19.15

Tabela 70 – Tabela de frequência para a variável FLG\_Claim\_Injured nos casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

044_Se existiram ou nao testemunhas		
FLG_Claim_Witness	Frequency	Percent
0	47	100.00

Tabela 71 – Tabela de frequência para a variável FLG\_Claim\_Witness nos casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

138_Indicador se ha Sinistros anteriores envolvendo o Segurado e Oficina em que seu veiculo foi reparado		
FLG_PC_Providers_CI	Frequency	Percent
0	44	93.62
1	3	6.38

Tabela 72 – Tabela de frequência para a variável FLG\_PC\_Providers\_CI nos casos onde  $FLG\_Susp\_Fraud = 1$  e  $FLG\_Conf\_Fraud = 0$ .

209_Indicador Cliente Veiculo Segurado escolheu a oficina		
FLG_VI_Chose_Garage	Frequency	Percent
0	47	100.00

Tabela 73 – Tabela de frequência para a variável FLG\_VI\_Chose\_Garage nos casos onde

*FLG\_Susp\_Fraud = 1 e FLG\_Conf\_Fraud = 0.*

210_Indicador se oficina associada ao Veiculo Segurado teve fraudes anteriores		
FLG_VI_Repair_PFraud	Frequency	Percent
0	47	100.00

Tabela 74 – Tabela de frequência para a variável FLG\_VI\_Repair\_PFraud nos casos onde

*FLG\_Susp\_Fraud = 1 e FLG\_Conf\_Fraud = 0.*

251_Indicador se ha Sinistros anteriores envolvendo a oficina do Veiculo 1		
FLG_PC_V1_Provider	Frequency	Percent
0	47	100.00

Tabela 75 – Tabela de frequência para a variável FLG\_PC\_V1\_Provider nos casos onde

*FLG\_Susp\_Fraud = 1 e FLG\_Conf\_Fraud = 0.*

Analysis Variable : VAL_Policy_Exposure 008_Valor Seguro total nesta Apolice (exposure)				
Mean	Std Dev	Minimum	Maximum	N
11769.65	22211.98	0	96654.57	44

Tabela 76 – Tabela de estatísticas descritivas para a variável VAL\_Policy\_Exposure nos casos onde

*FLG\_Susp\_Fraud = 1 e FLG\_Conf\_Fraud = 0.*

Variable	Mean	Std Dev	Minimum	Maximum	N
NUM_Antig_Alt_Policy	167.2179487	107.4427238	0	351	78
NUM_Antig_Policy	18.1025641	37.0831252	0	299	78
NUM_Days_Incept	5.9876543	10.7987196	0	81	81
NUM_V1_Age_FR	5	.	5	5	1
NUM_V1_ENGINE	.	.	.	.	0
NUM_V1_KM	.	.	.	.	0
NUM_V2_ENGINE	.	.	.	.	0
NUM_V2_KM	.	.	.	.	0
NUM_V2_Prior_Lines	0	0	0	0	3
NUM_VI_ENGINE	.	.	.	.	0
VAL_Prop_Exposure	1162.26	7454.22	0	60853.34	78
VAL_VI_INSURED_VALUE	.	.	.	.	0

Tabela 77 – Estatísticas descritivas para as variáveis numéricas nos casos onde *FLG\_Susp\_Fraud = 1 e*

*FLG\_Conf\_Fraud = 1.*

008_Valor Seguro total nesta Apolice (Em valores de percentagem do do valor mais elevado - 25%, 50% e 75%)		
Perc_Policy_Exposure	Frequency	Percent
25%	78	96.30
50%	1	1.23
75%	2	2.47

Tabela 78 – Tabela de frequência para a variável Perc\_Policy\_Exposure nos casos onde

*FLG\_Susp\_Fraud = 1 e FLG\_Conf\_Fraud = 1.*

034_Hora do Sinistro M:07h..12h, T:12h..20h, N:00h..07h, PN:20h..24h, S (sem info)		
COD_Claim_Time	Frequency	Percent
S	81	100.00

Tabela 79 – Tabela de frequência para a variável COD\_Claim\_Time nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 1.

039_Nut onde se deu o Acidente		
COD_Claim_Loc_Nuts	Frequency	Percent
Alentejo Litoral	1	1.23
Algarve	7	8.64
Alto Alentejo	2	2.47
Alto Trás-os-Montes	2	2.47
Ave	6	7.41
Baixo Alentejo	1	1.23
Baixo Mondego	4	4.94
Baixo Vouga	2	2.47
Beira Interior Norte	1	1.23
Beira Interior Sul	1	1.23
Cova da Beira	1	1.23
Cávado	3	3.70
Dão-Lafões	2	2.47
Entre Douro e Vouga	1	1.23
Grande Lisboa	16	19.75
Grande Porto	12	14.81
Minho-Lima	3	3.70
Médio Tejo	1	1.23
Oeste	1	1.23
Península de Setúbal	5	6.17
Pinhal Litoral	2	2.47
Tâmega	7	8.64

Tabela 80 – Tabela de frequência para a variável COD\_Claim\_Loc\_Nuts nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 1.

123_Nacionalidade do Segurado		
COD_NAT_CI	Frequency	Percent
EXTR	2	3.64
PORT	20	36.36
UNKN	33	60.00

Tabela 81 – Tabela de frequência para a variável COD\_NAT\_CI nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 1.

128_Sexo do Segurado		
COD_Gender_CI	Frequency	Percent
D	10	18.18
F	6	10.91
M	39	70.91

Tabela 82 – Tabela de frequência para a variável COD\_Gender\_CI nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 1.

COD_Occupation_CI	Frequency	Percent
Profissão detalhada	7	13.21
NULL	46	86.79

Tabela 83 – Tabela de frequência para a variável COD\_Occupation\_CI nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 1.

134_Natureza da Lesao no Segurado
--------------------------------------

COD_INJ_Nature_CI	Frequency	Percent

Tabela 84 – Tabela de frequência para a variável COD\_INJ\_Nature\_CI nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 1.

195_Marca e Modelo Veiculo Segurado		
COD_VI_MM	Frequency	Percent
AUDI	3	5.77
BMW	2	3.85
BOMBARDIER	1	1.92
DUCATI	1	1.92
FIAT	1	1.92
FORD	2	3.85
HONDA	4	7.69
HYUNDAI	1	1.92
JAGUAR	1	1.92
MERCEDES	8	15.38
MITSUBISHI	2	3.85
OPEL	5	9.62
PEUGEOT	4	7.69
PORSCHE	1	1.92
RENAULT	2	3.85
ROVER	2	3.85
SUZUKI	3	5.77
TOYOTA	4	7.69
VOLKSWAGEN	3	5.77
YAMAHA	2	3.85

Tabela 85 – Tabela de frequência para a variável COD\_VI\_MM nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 1.

196_Tipo do Veiculo Segurado		
COD_VI_TYPE	Frequency	Percent
OTH	1	1.92
VL	40	76.92
VM	11	21.15

Tabela 86 – Tabela de frequência para a variável COD\_VI\_TYPE nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 1.

243_Marca e Modelo Veiculo 1		
COD_V1_MM	Frequency	Percent
AUDI	3	15.79
BMW	2	10.53
CHRYSLER	1	5.26
FAMEL	1	5.26
FORD	1	5.26
LANCIA	1	5.26
MERCEDES	1	5.26
MINI	1	5.26
NISSAN	1	5.26
OPEL	2	10.53
ROVER	1	5.26
SMART	1	5.26
SUZUKI	1	5.26
VOLKSWAGEN	2	10.53

Tabela 87 – Tabela de frequência para a variável COD\_V1\_MM nos casos onde *FLG\_Susp\_Fraud* = 1 e

*FLG\_Conf\_Fraud* = 1.

244_Tipo do Veiculo 1		
COD_V1_TYPE	Frequency	Percent
VL	17	89.47
VM	2	10.53

Tabela 88 – Tabela de frequência para a variável COD\_V1\_TYPE nos casos onde *FLG\_Susp\_Fraud* = 1 e

*FLG\_Conf\_Fraud* = 1.

291_Marca e Modelo Veiculo 2		
COD_V2_MM	Frequency	Percent
PEUGEOT	1	33.33
RENAULT	1	33.33
SMART	1	33.33

Tabela 89 – Tabela de frequência para a variável COD\_V2\_MM nos casos onde *FLG\_Susp\_Fraud* = 1 e

*FLG\_Conf\_Fraud* = 1.

292_Tipo do Veiculo 2		
COD_V2_TYPE	Frequency	Percent
VL	3	100.00

Tabela 90 – Tabela de frequência para a variável COD\_V2\_TYPE nos casos onde *FLG\_Susp\_Fraud* = 1 e

*FLG\_Conf\_Fraud* = 1.

035_Se o Sinistro ocorreu ou nao durante um Final de Semana ou Feriado (Claim_dt) (tabela de feriados nacionais)		
FLG_Claim_Wkend	Frequency	Percent
0	58	71.60
1	23	28.40

Tabela 91 – Tabela de frequência para a variável FLG\_Claim\_Wkend nos casos onde *FLG\_Susp\_Fraud* = 1 e

*FLG\_Conf\_Fraud* = 1.

040_Se o Sinistro ocorreu no Estrangeiro ou em Portugal		
FLG_Claim_Overseas	Frequency	Percent
0	81	100.00

Tabela 92 – Tabela de frequência para a variável FLG\_Claim\_Overseas nos casos onde *FLG\_Susp\_Fraud* = 1 e

*FLG\_Conf\_Fraud* = 1.

041_Se o Tomador eh reclamante		
FLG_Claim_Insured	Frequency	Percent
0	81	100.00

Tabela 93 – Tabela de frequência para a variável FLG\_Claim\_Insured nos casos onde *FLG\_Susp\_Fraud* = 1 e

*FLG\_Conf\_Fraud* = 1.

042_Se existiu ou nao Intervencao das Autoridades		
FLG_Claim_Police_Rep	Frequency	Percent
0	81	100.00

Tabela 94 – Tabela de frequência para a variável FLG\_Claim\_Police\_Rep nos casos onde

*FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 1.

043_Se existiu ou nao Danos Corporais		
---------------------------------------	--	--

FLG_Claim_Injured	Frequency	Percent
0	71	87.65
1	10	12.35

Tabela 95 – Tabela de frequência para a variável FLG\_Claim\_Injured nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 1.

044_Se existiram ou nao testemunhas		
FLG_Claim_Witness	Frequency	Percent
0	81	100.00

Tabela 96 – Tabela de frequência para a variável FLG\_Claim\_Witness nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 1.

138_Indicador se ha Sinistros anteriores envolvendo o Segurado e Oficina em que seu veiculo foi reparado		
FLG_PC_Providers_CI	Frequency	Percent
0	76	93.83
1	5	6.17

Tabela 97 – Tabela de frequência para a variável FLG\_PC\_Providers\_CI nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 1.

209_Indicador Cliente Veiculo Segurado escolheu a oficina		
FLG_VI_Chose_Garage	Frequency	Percent
0	81	100.00

Tabela 98 – Tabela de frequência para a variável FLG\_VI\_Chose\_Garage nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 1.

210_Indicador se oficina associada ao Veiculo Segurado teve fraudes anteriores		
FLG_VI_Repair_PFraud	Frequency	Percent
0	81	100.00

Tabela 99 – Tabela de frequência para a variável FLG\_VI\_Repair\_PFraud nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 1.

251_Indicador se ha Sinistros anteriores envolvendo a oficina do Veiculo 1		
FLG_PC_V1_Provider	Frequency	Percent
0	81	100.00

Tabela 100 – Tabela de frequência para a variável FLG\_PC\_V1\_Provider nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 1.

Analysis Variable : VAL_Policy_Exposure 008_Valor Seguro total nesta Apolice (exposure)				
Mean	Std Dev	Minimum	Maximum	N
1606.00	6890.04	0	57331.83	77

Tabela 101 – Tabela de estatísticas descritivas para a variável VAL\_Policy\_Exposure nos casos onde *FLG\_Susp\_Fraud* = 1 e *FLG\_Conf\_Fraud* = 1.

Variavel	Nº Missing	% Total
NUM_Antig_Policy	38413	10.81%
NUM_Antig_Alt_Policy	20532	5.78%
VAL_Prop_Exposure	38413	10.81%
NUM_Days_Incept	25	0.01%

COD_Claim_Loc_Nuts	93523	26.31%
COD_NAT_CI	213005	59.92%
COD_Gender_CI	244177	68.69%
COD_Occupation_CI	351402	98.85%
COD_INJ_Nature_CI	355037	99.87%
COD_VI_MM	237087	66.69%
COD_VI_TYPE	244907	68.89%
VAL_VI_INSURED_VALUE	281815	79.28%
NUM_VI_ENGINE	265262	74.62%
NUM_V1_KM	345987	97.33%
NUM_V1_Age_FR	349325	98.27%
COD_V1_MM	317883	89.42%
COD_V1_TYPE	313782	88.27%
NUM_V1_ENGINE	350436	98.58%
NUM_V2_KM	354502	99.72%
COD_V2_MM	351127	98.77%
COD_V2_TYPE	350990	98.74%
NUM_V2_ENGINE	355072	99.88%
NUM_V2_Prior_Lines	350464	98.59%

Tabela 102 – Tabela de valores omissos para o total da amostra

Variavel	Nº Missing	% Segmento 1
NUM_Antig_Policy	38393	10.85%
NUM_Antig_Alt_Policy	20513	5.80%
VAL_Prop_Exposure	38393	10.85%
NUM_Days_Incept	25	0.01%
COD_Claim_Loc_Nuts	93054	26.30%
COD_NAT_CI	211890	59.89%
COD_Gender_CI	242978	68.67%
COD_Occupation_CI	349757	98.85%
COD_INJ_Nature_CI	353375	99.88%
COD_VI_MM	235936	66.68%
COD_VI_TYPE	243769	68.90%
VAL_VI_INSURED_VALUE	280386	79.25%
NUM_VI_ENGINE	263929	74.59%
NUM_V1_KM	344320	97.32%
NUM_V1_Age_FR	347669	98.26%
COD_V1_MM	316508	89.46%
COD_V1_TYPE	348772	98.57%
NUM_V1_ENGINE	312435	88.30%
NUM_V2_KM	352835	99.72%
COD_V2_MM	349497	98.78%



COD_V2_TYPE	349363	98.74%
NUM_V2_ENGINE	353406	99.88%
NUM_V2_Prior_Lines	348842	98.59%

Tabela 103 – Tabela de valores omissos para o segmento 1

Variavel	Nº Missing	% Segmento 2
NUM_Antig_Policy	14	0.91%
NUM_Antig_Alt_Policy	13	0.84%
VAL_Prop_Exposure	0	0.00%
NUM_Days_Incept	0	0.00%
COD_Claim_Loc_Nuts	469	30.47%
COD_NAT_CI	1078	70.05%
COD_Gender_CI	1162	75.50%
COD_Occupation_CI	1525	99.09%
COD_INJ_Nature_CI	1536	99.81%
COD_VI_MM	1107	71.93%
COD_VI_TYPE	1094	71.09%
VAL_VI_INSURED_VALUE	1301	84.54%
NUM_VI_ENGINE	1205	78.30%
NUM_V1_KM	1539	100.00%
NUM_V1_Age_FR	1531	99.48%
COD_V1_MM	1277	82.98%
COD_V1_TYPE	1249	81.16%
NUM_V1_ENGINE	1536	99.81%
NUM_V2_KM	1539	100.00%
COD_V2_MM	1508	97.99%
COD_V2_TYPE	1505	97.79%
NUM_V2_ENGINE	1538	99.94%
NUM_V2_Prior_Lines	1500	97.47%

Tabela 104 – Tabela de valores omissos para o segmento 2

Variavel	Nº Missing	% Segmento 3
NUM_Antig_Policy	3	6.38%
NUM_Antig_Alt_Policy	3	6.38%
VAL_Prop_Exposure	3	6.38%
NUM_Days_Incept	0	0.00%
COD_Claim_Loc_Nuts	0	0.00%
COD_NAT_CI	11	23.40%
COD_Gender_CI	11	23.40%
COD_Occupation_CI	46	97.87%
COD_INJ_Nature_CI	45	95.74%
COD_VI_MM	15	31.91%
COD_VI_TYPE	15	31.91%
VAL_VI_INSURED_VALUE	47	100.00%

NUM_VI_ENGINE	47	100.00%
NUM_V1_KM	47	100.00%
NUM_V1_Age_FR	45	95.74%
COD_V1_MM	36	76.60%
COD_V1_TYPE	36	76.60%
NUM_V1_ENGINE	47	100.00%
NUM_V2_KM	47	100.00%
COD_V2_MM	44	93.62%
COD_V2_TYPE	44	93.62%
NUM_V2_ENGINE	47	100.00%
NUM_V2_Prior_Lines	44	93.62%

Tabela 105 – Tabela de valores omissos para o segmento 3

Variavel	Nº Missing	% Segmento 4
NUM_Antig_Policy	3	3.70%
NUM_Antig_Alt_Policy	3	3.70%
VAL_Prop_Exposure	3	3.70%
NUM_Days_Incept	0	0.00%
COD_Claim_Loc_Nuts	0	0.00%
COD_NAT_CI	0	0.00%
COD_Gender_CI	26	32.10%
COD_Occupation_CI	74	91.36%
COD_INJ_Nature_CI	81	100.00%
COD_VI_MM	29	35.80%
COD_VI_TYPE	29	35.80%
VAL_VI_INSURED_VALUE	81	100.00%
NUM_VI_ENGINE	81	100.00%
NUM_V1_KM	81	100.00%
NUM_V1_Age_FR	80	98.77%
COD_V1_MM	62	76.54%
COD_V1_TYPE	62	76.54%
NUM_V1_ENGINE	81	100.00%
NUM_V2_KM	81	100.00%
COD_V2_MM	78	96.30%
COD_V2_TYPE	78	96.30%
NUM_V2_ENGINE	81	100.00%
NUM_V2_Prior_Lines	78	96.30%

Tabela 106 – Tabela de valores omissos para o segmento 4

### c) Matrizes

FLG_Claim_Wkend	Segmento 1	Segmento 2	Segmento 3	Segmento 4
Segmento 1		1	1	1
Segmento 2	1		1	1
Segmento 3	1	1		1
Segmento 4	1	1	1	

Matriz 8 – Relação da variável *FLG\_Claim\_Wkend* entre segmentos, sendo mais comum o facto de o acidente ocorrer durante a semana.

COD_Claim_Loc_Nuts	Segmento 1	Segmento 2	Segmento 3	Segmento 4
Segmento 1		1	1	1
Segmento 2	1		1	1
Segmento 3	1	1		1
Segmento 4	1	1	1	

Matriz 9 – Relação da variável *FLG\_Claim\_Wkend* entre segmentos, sendo mais comum o facto de os sinistros ocorrerem principalmente no grande Porto e grande Lisboa.

FLG_Claim_Overseas	Segmento 1	Segmento 2	Segmento 3	Segmento 4
Segmento 1		1	1	1
Segmento 2	1		1	1
Segmento 3	1	1		1
Segmento 4	1	1	1	

Matriz 10 – Relação da variável *FLG\_Claim\_Overseas* entre segmentos, sendo mais comum o facto de os sinistros em território português.

FLG_Claim_Insured	Segmento 1	Segmento 2	Segmento 3	Segmento 4
Segmento 1		1	1	1
Segmento 2	1		1	1
Segmento 3	1	1		1
Segmento 4	1	1	1	

Matriz 10 – Relação da variável *FLG\_Claim\_Insured* entre segmentos, sendo mais comum o facto de os sinistros não ter sido reclamado pelo tomador do seguro.

FLG_Claim_Police_Rep	Segmento 1	Segmento 2	Segmento 3	Segmento 4
Segmento 1		1	1	1
Segmento 2	1		1	1
Segmento 3	1	1		1
Segmento 4	1	1	1	

Matriz 11 – Relação da variável *FLG\_Claim\_Police\_Rep* entre segmentos, sendo mais comum o facto de os sinistros não terem obtido a presença das autoridades.

FLG_Claim_Injured	Segmento 1	Segmento 2	Segmento 3	Segmento 4
Segmento 1		1	1	1
Segmento 2	1		1	1
Segmento 3	1	1		1
Segmento 4	1	1	1	

Matriz 12 – Relação da variável *FLG\_Claim\_Injured* entre segmentos, sendo mais comum o facto de os sinistros não terem ocorrido lesões corporais decorrentes do mesmo.

COD_Gender_CI	Segmento 1	Segmento 2	Segmento 3	Segmento 4
Segmento 1		1	1	1
Segmento 2	1		1	1
Segmento 3	1	1		1
Segmento 4	1	1	1	

Matriz 13 – Relação da variável *COD\_Gender\_CI* entre segmentos, sendo mais comum o facto de os tomadores do seguro em questão serem do género masculino.

FLG_VI_Chose_Garage	Segmento 1	Segmento 2	Segmento 3	Segmento 4
Segmento 1		1	1	1
Segmento 2	1		1	1
Segmento 3	1	1		1
Segmento 4	1	1	1	

Matriz 14 – Relação da variável *FLG\_VI\_Chose\_Garage* entre segmentos, sendo mais comum o facto de os tomadores do seguro terem de sua vontade escolhido a oficina para as reparações dos seus veículos.

FLG_VI_Repair_PFraud	Segmento 1	Segmento 2	Segmento 3	Segmento 4
Segmento 1		1	1	1
Segmento 2	1		1	1
Segmento 3	1	1		1
Segmento 4	1	1	1	

Matriz 15 – Relação da variável *FLG\_VI\_Repair\_PFraud* entre segmentos, sendo mais comum o facto as oficinas seleccionadas pelos tomadores de seguros não terem sido identificadas por fraude anteriormente.

FLG_PC_V1_Provider	Segmento 1	Segmento 2	Segmento 3	Segmento 4
Segmento 1		1	1	1
Segmento 2	1		1	1
Segmento 3	1	1		1
Segmento 4	1	1	1	

Matriz 16 – Relação da variável *FLG\_PC\_V1\_Provider* entre segmentos, sendo mais comum o facto as oficinas seleccionadas pelos tomadores de seguros não terem sido seleccionadas em casos anteriores.